Le Trameur

CLA2T / SYLED - Université Sorbonne nouvelle Paris 3

Serge Fleury

Le Trameur
aka
Le Métier Textométrique

URL : http://www.tal.univ-paris3.fr/trameur
Téléchargement : http://www.tal.univ-paris3.fr/trameur
Contact : serge.fleury@univ-paris3.fr

Versions exécutables sous Windows disponibles en ligne
Documentation en ligne sur la page du projet

Manuel d'utilisation

version 12 *(et suivantes)*

mise à jour : printemps 2015

Sommaire

Table des Figures 7
2 Préambule 13
Trame et *Cadre* textométriques 15
3 Présentation générale 16
4 Installation 17
4 .1 Installation du Trameur 17
4 .2 Installation du Trameur (version console) 18
5 Interface du *Trameur* 19
6 Descriptif des icônes 20
7 Chargement de la base : mise au jour du Cadre et de la Trame 21
7.1 Chargement d'une nouvelle base 22
7.1.1 Paramétrage 22
7.1.2 Chargement 22
7.1.3 Etiquetage du texte 23
7.1.4 Inventaire typographique 24
7.1.5 Inventaire des parties 26
7.2 Importer une base 28
8 Les formats d'entrée (nouvelle base) 31
8.1 Format texte brut 31
8.2 Format dit « Lexico » 31
8.3 Format XML 31
9 Les dictionnaires 34
9.1 Les formes 34
9.2 Les lemmes 36
9.3 Les catégories 36
9.4 Les annotations importées 36
10 Accès à la Trame par le Cadre 37
10.1 Mise au jour de la Trame 37
10.2 Choix du niveau de l'annotation à afficher 40
10.3 Recherche de chaînes 43
10.4 Recherche « incrémentale » de chaînes 44
10.5 Visualisation des différentes annotations des items de la Trame 44
10.6 Edition et modification des items de la Trame 45
10.7 Le vocabulaire spécifique d'une partie 46
11 La carte des sections : représentation cartographique de la *Trame* 47
11.1 La carte des sections : une vue globale de la Trame 47
11.2 La carte des sections : une vue partielle de la Trame 50
11.3 La carte des sections : masquage de sections 52
12 Accès à la *Trame* par la carte des sections : principales fonctionnalités 53
12.1 Edition de section 53
12.2 Recherche de chaînes 55
12.3 Recherche « incrémentale » de chaînes 55
12.4 Visualisation des différentes annotations des items de la Trame 56
12.5 Edition et modification des items de la Trame 56
12.6 Le vocabulaire spécifique d'une section 57
12.7 Le vocabulaire spécifique d'une sélection de sections de la carte 57
12.8 Le vocabulaire spécifique des sections de la carte contenant un motif : recherche de cooccurrents 59
12.9 Le vocabulaire spécifique des sections de la carte ne contenant pas un motif 61
12.10 Fonctionnalités complémentaires sur la carte des sections 62
13 Concordances 63
13.1 Modes de déclenchement d'une concordance 64
13.2 Paramétrage d'une concordance 65
13.3 Editeur de concordance 67
14 Marquage des annotations de la *Trame* 68
14.1 Annotation « simple » 68
14.2 Annotation de type « relation » 73
15 Vocabulaire spécifique 75
16 Graphiques de ventilation 78
16.1 Ventilation par partie d'un item 78
16.2 Ventilation par partie d'une sélection d'items 80
16.3 Courbe d'accroissement du vocabulaire 81
16.5 Ventilation d'une sélection d'items associés à différents niveaux d'annotation 82
16.4 Ventilation d'une sélection d'items sur une sélection de parties 83
17 Segments répétés 85

18 Cooccurrences – Poly-Cooccurrences 87
18.1 Paramétrage et lancement des modules 87
18.2 Paramétrage de l'unité de contexte 88
18.2.1 Contexte défini par des délimiteurs 88
18.2.1 Contexte défini par le *Cadre* 89
18.3 Mise en œuvre du calcul 90
18.4 Cooccurrents et segments répétés 92
18.5 Sélection des chemins de cooccurrences à examiner 93
18.6 Fonctionnalités complémentaires sur les résultats produits 94
19 Extraction de patrons et graphes de mots 95
19.1 Extraction de patrons (recherche sur un niveau d'annotation) 95
19.2 Extraction de patrons (avec croisement d'annotations) 97
19.3 Actions sur les séquences extraites à partir d'un patron donné 98
19.4 Graphes de mots avec Le Trameur 101
19.5 Graphes de mots avec Pajek 102
20 Calculs textométriques complémentaires 105
20.1 Tableau lexical d'une partition 105
20.2 AFC : Analyse Factorielle des Correspondances 107
20.3 Intégration de modules R (phase 1) 110
Mise en œuvre des modules R dans le Trameur 111
Illustration sur la base « Etats finaux BROUILLONS » 112
21 Sélection. *Gestionnaire de sélection* 119
21.1 Sélection de positions sur la Trame 119
21.1.1 Sélection directe à partir de la *Trame* 120
21.1.2 Sélectionner dans les dictionnaires 121
21.2 Traitements d'une sélection dans le Gestionnaire de sélection 122
21.3 Opérations sur sélections : concordance, ventilation, modification annotation, nouvelle annotation 123
21.3.1 Concordance 123
21.3.2 Ventilation 125
21.3.3 Modification des annotations sur les items sélectionnés 126
21.3.4 Ajout d'une nouvelle annotation sur les items sélectionnés 128
21.4 Méthodes de génération de sélection via le Gestionnaire 130
21.4.1 Génération et sélection d'une liste de formes 130
21.4.2 Génération et sélection d'une liste de segments 131
21.4.3 Génération et sélection multicritère 131
21.4.4 Génération d'une liste d'items spécifiques 132
21.5 Sélection de parties (comme ensembles de sélection de positions sur la Trame) 134
21.5.1 Calcul de segments 136
21.5.2 Calcul de spécificité 137
21.5.3 Extraction de patron 138
21.5.4 Calcul de cooccurrents 139
21.5.5 Calcul de cooccurrents sur corpus chronologique 141
21.6 Gestion des sélections : sauvegarde et chargement 147
21.6.1 Sauvegarde d'une Sélection 148
21.6.2 Chargement d'une Sélection 152
22 Importation d'annotation sur une *Trame* 154
23 Ajout/modification d'annotation sur une *Trame* via les dictionnaires 160
23.1 Ajout d'annotation 161
23.2 Modification d'annotation 162
24 Des nuages de d'annotations sur la *Trame* 163
25 Relations de dépendance entre les items de *Trame* (via leurs annotations) 165
25.1 Extrait de la Trame : « un jour, un petit ange de sapin de … » 166
25.2 Recherche de relations 167
25.2.1 Sélection des relations à afficher 167
25.2.2 Graphe des relations 168
25.3 Sélection des relations avec filtrage sur le nœud source de la relation 168
25.4 Sélection des relations avec filtrage sur la position du dépendant 171
25.5 Actions sur les nœuds du graphe : retour en contexte 171
25.6 Rechercher dans un graphe de dépendance 175
25.7 Affichage de relation en contexte 177
25.8 Recherche de relation en contexte 179
25.9 Graphe de relations en contexte 182
25.10 Visualisation de toutes les relations dans un contexte donné 183
25.11 Recherche de collocation : spécificités sur relation 185
25.12 Recherche source-cible d'une relation 187
26 Le *Cadre* : visualisation, réorganisation 190
27 Exporter une base textométrique 193

28 Exportation du *Cadre* et de la *Trame* 199
28.1 Fichier de travail 199
28.2 Fichiers d'export 199
29 Le rapport : enregistrement des résultats produits 206
29.1 Ajouter un élément au rapport 206
29.2 Ajouter un élément externe au rapport 207
29.2 Enregistrer le rapport 208
30 Fonctionnalités complémentaires 211
30.1 Graphes d'annotation 211
30.1.1 Intégration dans le *Trameur* d'un module d'importation d'un graphe d'annotations sur la *Trame* 211
30.1.1 Génération de graphes d'annotations 214
30.1.3 Exploitation des graphes d'annotation 218
30.2 Base textométrique de textes alignés 219
30.2.1 Mise au jour d'alignement dans la carte des sections 219
30.2.2 Exploration d'une base textométrique de textes alignés 222
30.2.3 Outils et méthodes d'exploration d'un alignement 240
30.3 Annotation de données orales 245
30.4 Module petiMoteur 247
31 Le Trameur en mode console 248
31.1 Fichier de travail : 248
31.2 Le Cadre : 249
31.3 La Trame : 250
31.4 La Trame « annotée » (si treetagger est invoqué) 252
32 Bibliographie 254
33 Glossaire 256

Table des Figures

Figure 1 : *Le Trameur* Fenêtre principale 19
Figure 2 : *Le Trameur* Onglet PARAM 22
Figure 3 : *Le Trameur* Onglet CADRE 23
Figure 4 : *Le Trameur* Onglet PARAM : choix de langue pour Treetagger 23
Figure 5a : *Le Trameur* Inventaire Typographique 24
Figure 5b : *Le Trameur* Inventaire Typographiquesur petitMoteur (1) 25
Figure 5c : *Le Trameur* Inventaire Typographiquesur petitMoteur (2) 26
Figure 5d : *Le Trameur* Inventaire des parties 26
Figure 5e : *Le Trameur* Inventaire des parties sur Gromoteur 27
Figure 6 : *Le Trameur* Onglet FORME 34
Figure 7 : *Le Trameur* Actions sur Items sélectionnés à partir des dictionnaires 35
Figure 8 : *Le Trameur* Onglet FORME 36
Figure 9 : *Le Trameur* Onglet CADRE, base chargée 38
Figure 10 : *Le Trameur* Onglet CADRE, *Cadre* généré 38
Figure 11 : *Le Trameur* Onglet CADRE, contenu textuel d'une partie (les mots) 39
Figure 12 : *Le Trameur* Onglet CADRE, contenu textuel d'une partie (les lemmes) 39
Figure 13 : *Le Trameur* Onglet CADRE, chargement d'une représentation arborescente d'une base au format XML (1) 40
Figure 14 : *Le Trameur* Onglet CADRE, chargement d'une représentation arborescente d'une base au format XML (2) 40
Figure 15 : *Le Trameur* Affichage d'une partie du CADRE (annotation affichée = 1) 41
Figure 16 : *Le Trameur* Affichage d'une partie du CADRE (annotation affichée = 2) 42
Figure 17 : *Le Trameur* Affichage d'une section (annotation affichée = 3) 42
Figure 18 : *Le Trameur* Onglet CADRE, recherche de chaînes 43
Figure 19 : *Le Trameur* Onglet CADRE, recherche incrémentale de chaînes 44
Figure 20 : *Le Trameur* Onglet CADRE, visualisation détaillée d'un item 44
Figure 21 : *Le Trameur* Onglet CADRE, édition d'un item 45
Figure 22 : *Le Trameur* Onglet CADRE, modification d'un item 45
Figure 23 : *Le Trameur* Onglet CADRE, le vocabulaire spécifique d'une partie 46
Figure 24 : *Le Trameur* Onglet SECTION, texte brut avec caractère « délimiteur de sections » 47
Figure 25 : *Le Trameur* Onglet SECTION, texte structuré avec caractère « délimiteur de sections » 48
Figure 26 : *Le Trameur* Onglet SECTION, construction de la carte des sections (3) 48
Figure 27 : *Le Trameur* Onglet SECTION, édition de la *Trame* par une section 49
Figure 28 : *Le Trameur* Onglet SECTION, paramétrage de l'affichage 49
Figure 29 : *Le Trameur* Onglet SECTION, texte TEI sans caractère « délimiteur de sections » explicite 50
Figure 30 : *Le Trameur* Onglet SECTION, sections définies implicitement par un élement du *Cadre* (avec affichage d'une autre partie : lg) 51
Figure 31 : *Le Trameur* Onglet SECTION, sections définies implicitement par un élement du *Cadre* (avec affichage d'une autre partie : body) 51
Figure 32 : *Le Trameur* Onglet SECTION, masquage de sections (1) 52
Figure 33 : *Le Trameur* Onglet SECTION, masquage de sections (2) 52
Figure 34 : *Le Trameur* Onglet SECTION, édition de section 53
Figure 35 : *Le Trameur* Onglet SECTION, édition de section, sélection des annotations 54
Figure 36 : *Le Trameur* Onglet SECTION, recherche de chaînes 55
Figure 37 : *Le Trameur* Onglet SECTION, recherche de chaînes 56
Figure 38 : *Le Trameur* Onglet SECTION, les mots spécifiques d'une section 57
Figure 39 : *Le Trameur* Onglet SECTION, les mots spécifiques d'une sélection de sections 58
Figure 40 : *Le Trameur* Onglet SECTION, items spécifiques des sections contenant un motif (cooccurrents) 60
Figure 41 : *Le Trameur* Onglet SECTION, items spécifiques des sections ne contenant pas un motif 61
Figure 42 : *Le Trameur* Onglet CONCORDANCE, affichage édition + affichage tri 63
Figure 43 : *Le Trameur* Onglet CONCORDANCE (forme) 64
Figure 44 : *Le Trameur* Onglet CONCORDANCE (lemme) 64
Figure 45 : *Le Trameur* Onglet CONCORDANCE (catégorie) 65
Figure 46 : *Le Trameur* Onglet CONCORDANCE (forme) 66
Figure 47 : *Le Trameur* Onglet CONCORDANCE (rechargement) 66
Figure 48 : *Le Trameur* Onglet CONCORDANCE, édition et modification d'un item pivot 67
Figure 49 : *Le Trameur* Editeur CADRE 68
Figure 50 : *Le Trameur* Editeur SECTION 69

Figure 51 : *Le Trameur* Editeur CONCORDANCE 69
Figure 52 : *Le Trameur* Sélection des annotations à colorier (1) 70
Figure 53 : *Le Trameur* Sélection des annotations à colorier (2) 70
Figure 54 : *Le Trameur* Marquage des annotations sur les items édités (1) 71
Figure 55 : *Le Trameur* Marquage des annotations sur les items édités (2) 71
Figure 56 : *Le Trameur* Marquage des annotations sur les items édités (3) 71
Figure 57 : *Le Trameur* Sélection d'une nouvelle annotation à colorier 72
Figure 58 : *Le Trameur* Marquage des annotations sur les items édités 72
Figure 59 : *Le Trameur* Edition d'une annotation de type « relation » 73
Figure 60 : *Le Trameur* Sélection d'une relation à colorier 74
Figure 61 : *Le Trameur* Marquage des annotations de type « relation » sur les items édités 74
Figure 62 : *Le Trameur* Sélection d'annotation pour calcul de spécificités 75
Figure 63 : *Le Trameur* Opérations disponibles sur les items issus d'un calcul de spécificités 76
Figure 64 : *Le Trameur* Opérations disponibles sur les items triables issus d'un calcul de spécificités 77
Figure 65 : *Le Trameur* Onglet VENTILATION paramétrage d'une ventilation 78
Figure 66 : *Le Trameur* Onglet VENTILATION affichage d'une ventilation (mode ligne) 79
Figure 67 : *Le Trameur* Onglet VENTILATION mode regrexp 79
Figure 68 : *Le Trameur* Ventilation d'une sélection d'items 80
Figure 69 : *Le Trameur* Onglet VENTILATION, courbe d'accroissement du vocabulaire 81
Figure 70 : *Le Trameur* Onglet VENTILATION, ventilation croisée d'annotations 82
Figure 71 : *Le Trameur* Onglet VENTILATION, ventilation d'une sélection d'items (1) 83
Figure 72 : *Le Trameur* Onglet VENTILATION, ventilation d'une sélection d'items (2) 83
Figure 73 : *Le Trameur* Onglet VENTILATION, ventilation d'une sélection d'items (3) 84
Figure 74 : *Le Trameur* Onglet SEGMENT 85
Figure 75 : *Le Trameur* Onglet SEGMENT 86
Figure 76 : *Le Trameur* Onglet COOC, importation d'une stop-liste 90
Figure 77 : *Le Trameur* Onglet COOC, édition de la stop-liste 91
Figure 78 : *Le Trameur* Onglet COOC, étape n°1 : module Cooccurrents 91
Figure 79 : *Le Trameur* Onglet COOC, étape n°2 : module Polycooccurrents 92
Figure 80 : *Le Trameur* Onglet COOC, un pôle de type SR 92
Figure 81 : *Le Trameur* Onglet COOC, Coocs + SR 93
Figure 82 : *Le Trameur* Onglet COOC, Filtrage des polycoocs 93
Figure 83 : *Le Trameur* onglet PATRON 95
Figure 84 : *Le Trameur* zone d'enregistrement d'un patron 96
Figure 85 : *Le Trameur* écriture d'un patron 96
Figure 86 : *Le Trameur* liste des séquences associées au patron 96
Figure 87 : *Le Trameur* écriture et paramétrage d'un patron 96
Figure 88 : *Le Trameur* liste des séquences associées au patron 97
Figure 89 : *Le Trameur* Extraction de patron (croisement d'annotation) 97
Figure 90 : *Le Trameur* Extraction de patron (croisement d'annotation) 98
Figure 91 : *Le Trameur* Extraction de patron (croisement d'annotation) 98
Figure 92 : *Le Trameur* Paramétrage de l'extraction du patron visé et de l'affichage des séquences extraites 99
Figure 93 : *Le Trameur* Affichage des séquences extraites sur la base du patron visé 100
Figure 94 : *Le Trameur* Ventilation-partie d'une liste de séquences sélectionnées 100
Figure 95 : *Le Trameur* Concordance d'une liste de séquences sélectionnées 100
Figure 96 : *Le Trameur* Ventilation-section d'une liste de séquences sélectionnées 100
Figure 97 : *Le Trameur* génération du graphe de mots complet associé à un patron 101
Figure 98 : *Le Trameur* génération du graphe de mots associé à un patron et contenant un motif 102
Figure 99 : *Le Trameur* export au format *Pajek* 103
Figure 100 : *Le Trameur* export NOM ADJ au format *Pajek* sur Duchn 103
Figure 101 : Le réseau NOM ADJ de Duchn *via Pajek* 104
Figure 102 : Calculs textométriques complémentaires 105
Figure 103 : Sélection du niveau d'annotation pour calculs lexicométriques complémentaires 105
Figure 104 : Paramétrage TFG ; choix des paramètres de calcul, de l'annotation et de la partition 106
Figure 105 : Tableau lexical (annotation n°1 : forme) 106
Figure 106 : Tableau lexical (annotation n°3 : catégorie) 107
Figure 107 : AFC 108
Figure 108 : AFC (position) 108
Figure 109 : AFC (sélection) 108
Figure 110 : AFC (affichage des mots spécifiques de la partie) 109

Figure 111 : Les facteurs de l'AFC 109
Figure 112 : PCLC d'une partition 110
Figure 113 : Modules R (1) 111
Figure 114 : Modules R (2) 111
Figure 115 : AFC via R 112
Figure 116 : ACP via R 113
Figure 117 : CAH via R 114
Figure 118 : Sélection des items pour les modules R 115
Figure 119 : Items sélectionnés dans la *Gestionnaire de Sélection* 116
Figure 120 : Paramétrage R pour travailler avec le *Gestionnaire de Sélection* 116
Figure 121 : AFC via R sur le contenu du *Gestionnaire de Sélection* 117
Figure 122 : CAH via R sur le contenu du *Gestionnaire de Sélection* 118
Figure 123 : *Le Trameur* onglet SELECTION 119
Figure 124 : *Le Trameur* sélection sur la *Trame* (onglet CADRE) 120
Figure 125 : *Le Trameur* sélection sur la *Trame* (onglet VENTILATION) 120
Figure 126 : *Le Trameur* sélection sur une concordance (onglet CONCORDANCE) 121
Figure 127 : *Le Trameur* sélection sur dictionnaires (onglets FORME, LEMME, CATEGORIE) 121
Figure 128 : *Le Trameur Gestionnaire de Sélection* 122
Figure 129 : *Le Trameur Gestionnaire de Sélection* (positions stockées dans le *Gestionnaire de Sélection*) 123
Figure 130 : *Le Trameur* sélection directe de position dans l'éditeur 124
Figure 131 : *Le Trameur Gestionnaire de Sélection* : les positions enregistrées 124
Figure 132 : *Le Trameur Gestionnaire de Sélection* : sélection parmi les positions enregistrées 125
Figure 133 : *Le Trameur* concordance d'une sélection de positions 125
Figure 134 : *Le Trameur* ventilation d'une sélection de positions 126
Figure 135 : *Le Trameur* modification des annotations d'une sélection de positions 127
Figure 136 : *Le Trameur* modification des annotations d'une sélection de positions 127
Figure 137 : *Le Trameur* modification des annotations d'une sélection de positions 128
Figure 138 : *Le Trameur* insertion d'une nouvelle annotation sur une sélection de positions 128
Figure 139 : *Le Trameur* insertion d'une nouvelle annotation sur une sélection de positions 129
Figure 140 : *Le Trameur* insertion d'une nouvelle annotation sur une sélection de positions 129
Figure 141 : *Le Trameur* insertion d'une nouvelle annotation sur une sélection de positions 129
Figure 142 : *Le Trameur* génération d'une liste de formes à sélectionner 130
Figure 143 : *Le Trameur* sélection d'une liste de formes 130
Figure 144 : *Le Trameur* génération d'une liste de segments à sélectionner 131
Figure 145 : *Le Trameur* génération d'une liste d'items par équation de recherche 132
Figure 146 : *Le Trameur* sélection de parties 135
Figure 147 : *Le Trameur* calcul de segment sur une sélection de parties 136
Figure 148 : *Le Trameur* calcul de segment sur une sélection de parties 137
Figure 149 : *Le Trameur* calcul de spécificités sur une sélection de parties 137
Figure 150 : *Le Trameur* extraction de patron sur une partie sélectionnée 138
Figure 151 : *Le Trameur* extraction de patron sur une partie sélectionnée 139
Figure 152 : *Le Trameur* cooccurrents/polycooccurrents sur une partie sélectionnée (1) 139
Figure 153 : *Le Trameur* cooccurrents/polycooccurrents sur une partie sélectionnée (2) 140
Figure 154 : *Le Trameur* cooccurrents/polycooccurrents sur une partie sélectionnée (3) 141
Figure 155 : Cooccurrents, approche locale / globale / chronologique 142
Figure 156 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (1) 142
Figure 157 : Cooccurrents, approche locale 143
Figure 158 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (2) 143
Figure 159 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (3) 144
Figure 160 : Cooccurrents, approche globale 145
Figure 161 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (4) 145
Figure 162 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (5) 146
Figure 163 : Cooccurrents, approche chronologique 147
Figure 164 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (6) 147
Figure 165 : *Le Trameur* Gestion des sélections : sauvegarde et chargement 148
Figure 166 : *Le Trameur* sauvegarde d'une sélection (1) 148
Figure 167 : *Le Trameur* sauvegarde d'une sélection (2) 148
Figure 168 : *Le Trameur* sauvegarde d'une sélection (3) 149
Figure 169 : *Le Trameur* arborescence XML d'une sélection 149
Figure 170 : *Le Trameur* chargement d'une sélection (1) 152

Figure 171 : *Le Trameur* chargement d'une sélection (2) 152
Figure 172 : *Le Trameur* activation d'une sélection nommée (1) 153
Figure 173 : *Le Trameur* activation d'une sélection nommée (2) 153
Figure 174 : *Le Trameur* inhibation d'une sélection nommée 153
Figure 175 : *Le Trameur* Importation d'annotation, nommage et sélection du niveau d'annotation 157
Figure 176 : *Le Trameur* correction d'item (annotation importée) 158
Figure 177 : *Le Trameur* concordance d'item (annotation importée) 158
Figure 178 : *Le Trameur* ventilation d'item (annotation importée) 159
Figure 179 : *Le Trameur* édition de la *Trame* par les formes 160
Figure 180 : *Le Trameur* ajout d'annotation sur la *Trame* via une forme sélectionnée 161
Figure 181 : *Le Trameur* modification d'annotation sur la *Trame* via une forme sélectionnée 162
Figure 182 : *Le Trameur* un nuage d'annotations sur la *Trame* 163
Figure 183 : *Le Trameur* un nuage d'annotations sur la *Trame* avec contextes 163
Figure 184 : *Le Trameur* paramétrage du nuage d'annotations 164
Figure 185 : *Le Trameur* recherche de relations de dépendance via l'onglet GRAPHE 167
Figure 186 : *Le Trameur* paramétrage de la recherche de dépendance sur la *Trame* 167
Figure 187 : *Le Trameur* affichage du graphe des dépendances 168
Figure 188 : *Le Trameur* paramétrage de la recherche de dépendance avec filtrage 169
Figure 189 : *Le Trameur* affichage du graphe des dépendances (1) 169
Figure 190 : *Le Trameur* affichage du graphe des dépendances (2) 170
Figure 191 : *Le Trameur* affichage du graphe des dépendances (3) 170
Figure 192 : *Le Trameur* paramétrage de la recherche de dépendance avec filtrage 172
Figure 193 : *Le Trameur* affichage du graphe des dépendances 172
Figure 194 : *Le Trameur* affichage de la concordance sur un des nœuds du graphe des dépendances 172
Figure 195 : *Le Trameur* affichage de la concordance sur tous les nœuds du graphe des dépendances 173
Figure 196 : *Le Trameur* Concordance sur tous les nœuds du graphe des dépendances (1) 173
Figure 197 : *Le Trameur* Concordance sur tous les nœuds du graphe des dépendances (2) 173
Figure 198 : *Le Trameur* Coloration d'annotation dans le concordancier (1) 174
Figure 199 : *Le Trameur* Coloration d'annotation dans le concordancier (2) 174
Figure 200 : *Le Trameur* Graphe de la relation OBL 175
Figure 201 : *Le Trameur* Recherche dans un graphe de relation (1) 176
Figure 202 : *Le Trameur* Recherche dans un graphe de relation (2) 176
Figure 203 : *Le Trameur* Recherche dans un graphe de relation (3) 177
Figure 204 : *Le Trameur* accès aux dépendances via la *Trame* 178
Figure 205 : *Le Trameur* paramétrage de l'affichage des dépendances sur un item de la *Trame* (1) 178
Figure 206 : *Le Trameur* paramétrage de l'affichage des dépendances sur un item de la *Trame* (2) 178
Figure 207 : *Le Trameur* affichage de la dépendance sur la *Trame* (1) 179
Figure 208 : *Le Trameur* affichage de la dépendance via un graphe (2) 179
Figure 209 : *Le Trameur* recherche de dépendance en contexte sur un item de la *Trame* 179
Figure 210 : *Le Trameur* paramétrage de la recherche de dépendance sur un item de la *Trame* (1) 180
Figure 211 : *Le Trameur* affichage des dépendances via un graphe (1) 180
Figure 212 : *Le Trameur* affichage des dépendances en contexte sur la *Trame* 180
Figure 213 : *Le Trameur* paramétrage de la recherche de dépendance sur un item de la *Trame* (2) 181
Figure 214 : *Le Trameur* affichage des dépendances via un graphe (2) 181
Figure 215 : *Le Trameur* affichage d'un graphe de dépendances en contexte (1) 182
Figure 216 : *Le Trameur* affichage d'un graphe de dépendances en contexte (2) 182
Figure 217 : *Le Trameur* affichage global des dépendances en contexte (1) 183
Figure 218 : *Le Trameur* affichage global des dépendances en contexte (2) 183
Figure 219 : *Le Trameur* affichage global des dépendances en contexte (3) 184
Figure 220 : *Le Trameur* affichage global des dépendances en contexte (4) 184
Figure 221 : *Le Trameur* Recherche de collocation (1) 185
Figure 222 : *Le Trameur* Recherche de collocation (2) 186
Figure 223 : *Le Trameur* Recherche de collocation (3) 186
Figure 224 : *Le Trameur* liste de relations 187
Figure 225 : *Le Trameur* liste de relations (cible) 187
Figure 226 : *Le Trameur* liste de POS en relation 188
Figure 227 : *Le Trameur* les dépendants d'un item (forme, lemme ou catégorie) 188
Figure 228 : *Le Trameur* les dépendants d'une liste d'items (forme, lemme ou catégorie) 189
Figure 229 : *Le Trameur* retour en contexte (les dépendants d'un item) 189
Figure 230 : *Le Trameur* le *Cadre* (visualisation, réorganisation) 190

Figure 231 : *Le Trameur* Chargement d'une partition à réorganiser 191
Figure 232 : *Le Trameur* Réorganisation d'une partition 191
Figure 233 : *Le Trameur* Courbe d'accroissement sur une partition initiale et la même partition tronquée et réorganisée 192
Figure 234 : Ajouter un élément au rapport 206
Figure 235 : Rapport en cours dans l'onglet RAPPORT 207
Figure 236 : Ajouter un élément externe au rapport 208
Figure 237 : Rapport enregistré 208
Figure 238 : Index du rapport enregistré 209
Figure 239 : Elément du rapport enregistré 210
Figure 240 : Importation d'annotation+graphe 213
Figure 241 : Chargement de la base, accès à la *Trame* par le *Cadre* 214
Figure 242 : Sélection des items constituant un nœud du graphe 215
Figure 243 : Construction du nœud 215
Figure 244 : Insertion d'un label sur le nœud 215
Figure 245 : Le nœud a un label 215
Figure 246 : De nouveaux nœuds… 216
Figure 247 : La source d'un arc 216
Figure 248 : La cible de l'arc... l'arc est défini 216
Figure 249 : Ajout d'un label sur l'arc 217
Figure 250 : L'arc a son label 217
Figure 251 : Graphe complet construit 217
Figure 252 : Sauvegarde du graphe 218
Figure 253 : Alignement, partition tu (1) 220
Figure 254 : Alignement, partition tu (2) 220
Figure 255 : Alignement, partition volet (1) 221
Figure 256 : Alignement, partition volet (2) 221
Figure 257 : fichier 1, volet français (extrait) 222
Figure 258 : fichier 1, volet anglais (extrait) 222
Figure 259 : Edition d'un alignement dans mkAlign 223
Figure 260 : Préparation d'une base textométrique de textes bilingues alignés 223
Figure 261 : Carte des sections d'un alignement dans le Trameur 224
Figure 262 : Carte des sections d'un alignement dans mkAlign 224
Figure 263 : Projection d'une forme sur carte des sections dans mkAlign 224
Figure 264 : Projection d'une forme sur la carte des sections dans Le Trameur 224
Figure 265 : Paramétrage d'un bitexte 225
Figure 266 : Mise au jour du bitexte via la carte des sections 225
Figure 267 : Spécificités volet SOURCE de l'alignement 226
Figure 268 : Spécificités volet CIBLE de l'alignement 226
Figure 269 : Spécificités des volets SOURCE et CIBLE via mkAlign 227
Figure 270 : Base textométrique de 2 traductions alignées 227
Figure 271 : Carte des sections d'un alignement de traductions dans le Trameur 228
Figure 272 : Projection d'une forme sur les 2 traductions alignées 228
Figure 273 : Spécificités sur volet SOURCE (TRADUCTION2) : sections contenant le motif "vieil" 229
Figure 274 : Spécificités sur volet CIBLE (TRADUCTION1) : 229
sections MIROIR des sections contenant le motif "vieil" 229
Figure 275 : Alignement TMX dans un navigateur (avec feuille de styles) 230
Figure 276 : Alignement TMX (code source) 230
Figure 277 : Alignement TMX , arbre des éléments 231
Figure 278 : Carte des sections d'un alignement TMX avec le Trameur 231
Figure 279 : 1 colonne = 1 volet = 1 langue 232
Figure 280 : Distribution d'une forme sur la carte de l'alignement 232
Figure 281 : Contraste de volets (2 à 2) 233
Figure 282 : Contraste de 2 volets (FR2, FR1) 234
Figure 283 : Spécificités volet SOURCE (FR2) 234
Figure 284 : Spécificités volet CIBLE (FR1) 235
Figure 285 : Contraste de 2 volets (FR2, EN) 235
Figure 286 : Spécificités volet SOURCE (FR2) 236
Figure 287 : Spécificités volet CIBLE (EN) 236
Figure 288 : Base textométrique multiannotée d'un alignement de textes bilingues 238

Figure 289 : Calculs contrastifs sur une base textométrique d'alignements 239
Figure 290 : Carte de sections d'un alignement multi-volets 240
Figure 291 : Edition d'un alignement multi-volets 241
Figure 292 : Opérations dans une section multi-volets 242
Figure 293 : Graphe de relations de dépendances dans une section multi-volets (1) 243
Figure 294 : Graphe de relations de dépendances dans une section multi-volets (2) 244
Figure 295 : Graphe de relations de dépendances dans une section multi-volets (3) 244
Figure 296 : CFPP2000 (1) 245
Figure 297 : CFPP2000 (2) 246
Figure 298 : Le petiMoteur 247
Figure 299 : *Le Trameur* en mode console 248

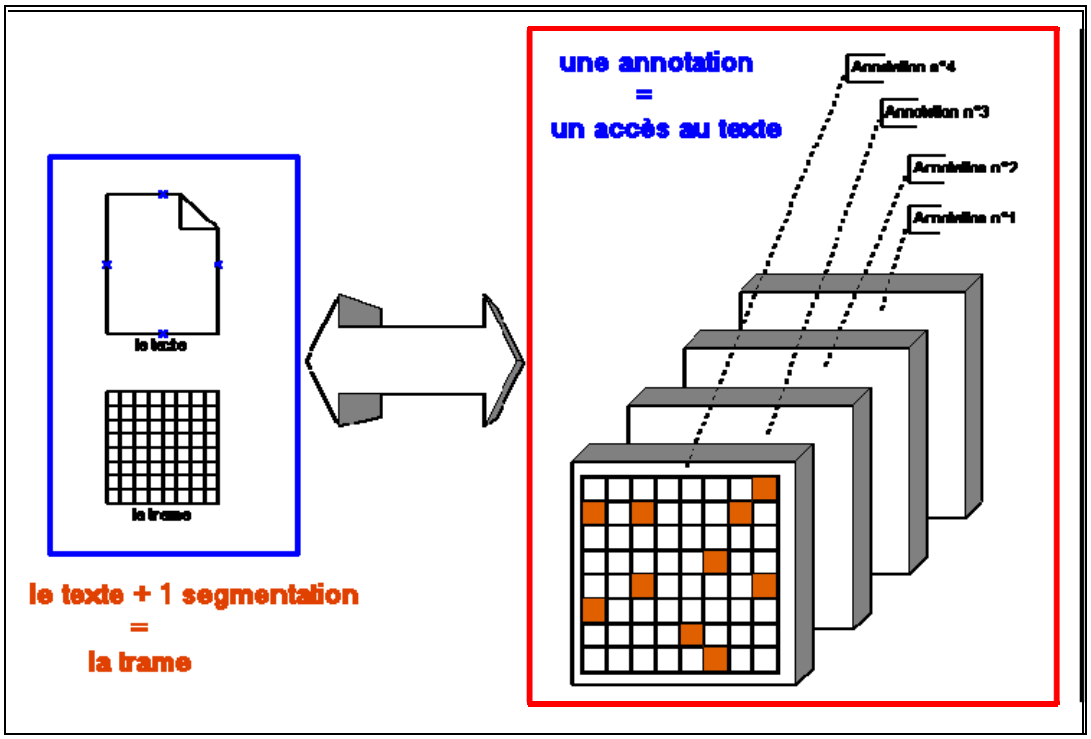
2 Préambule

L'objectif principal de la textométrie est de compter des éléments (des *contenus* textuels) dans des ensembles (des *contenants* regroupant des unités élémentaires d'un texte ou des zones de texte couvrant un certain nombre ou un certain type d'unités élémentaires).

- Les *contenus* se réalisent sous la forme de ressources textuelles (une séquence de caractères organisée en phrases, en paragraphes etc.).

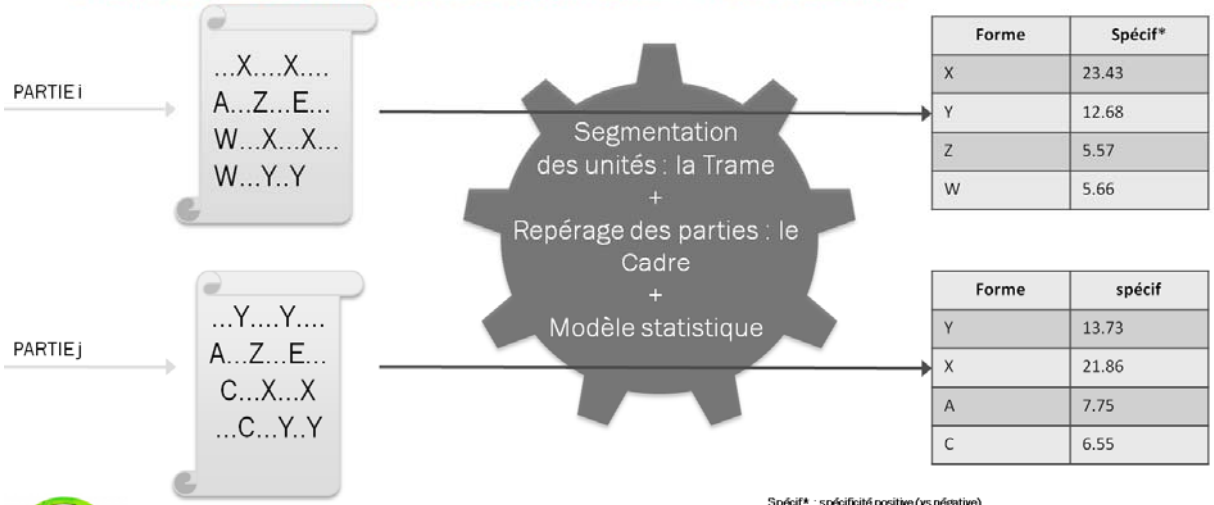
- Les *contenants* existent sous la forme de système de masques ou de calques que l'on peut définir sur les contenus. Il s'agit de systèmes d'annotations que l'on peut définir sur tout ou partie des zones textuelles, ces annotations constituant en retour des accès sur les parties textuelles qu'elles définissent (le marquage des phrases ou des paragraphes étant un exemple d'annotation particulier pour décrire un certain niveau de la structure du texte).

Le processus de comptage nécessite au préalable d'identifier les *contenus* et les *contenants*. Ce préalable consiste à expliciter une *segmentation* du texte conduisant à la mise au jour d'une ***Trame*** sur laquelle des annotations pourront se greffer ultérieurement.



L'objectif de la textométrie est de compter des éléments (des *contenus* textuels) dans des ensembles (des *contenants)* regroupant des unités élémentaires d'un texte ou des zones de texte couvrant un certain nombre ou un certain type d'unités élémentaires

« Classiquement », la textométrie construit des tableaux lexicaux à partir des données à analyser sur la base de processus de segmentation + modèle statistique



Forme	Spécif*
X	23.43
Y	12.68
Z	5.57
W	5.66

Forme	spécif
Y	13.73
X	21.86
A	7.75
C	6.55

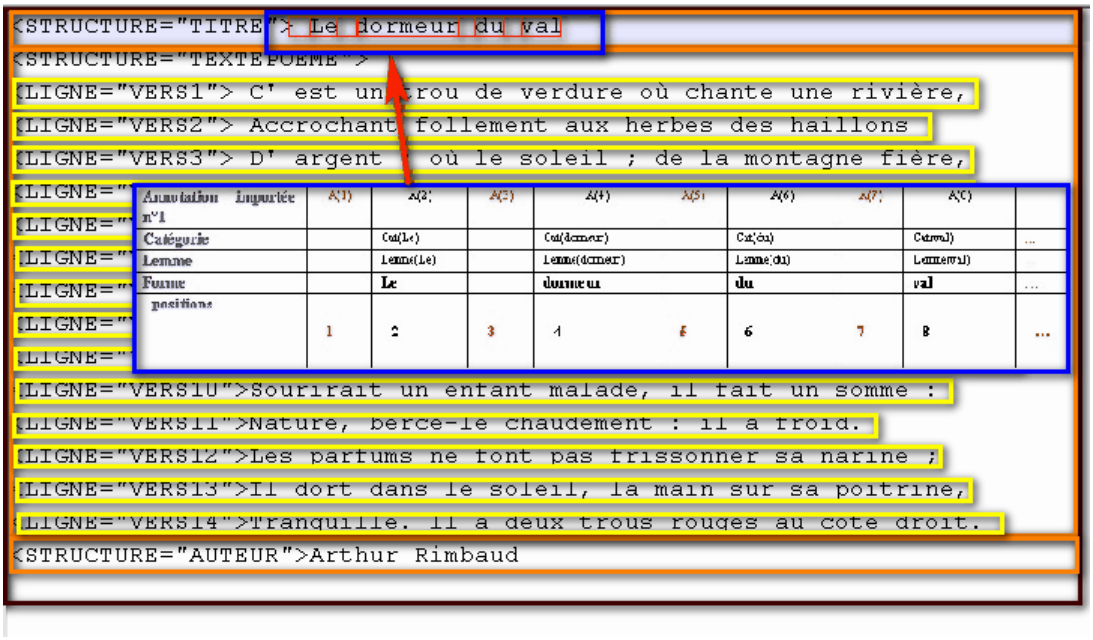
Dans le cas du Trameur, aucun tableau lexical n'est disponible directement (calcul dynamique en cas de besoin) : les unités ne sont pas « figées »

Spécif* : spécificité positive (vs négative)
pour un seuil de spécificité fixé, une forme i et une partie j données, la forme i est dite spécifique positive de la partie j si sa sous-fréquence est "anormalement élevée" dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou supérieures à la sous-fréquence constatée est inférieure au seuil fixé au départ.

Trame et *Cadre* textométriques

*A partir d'un texte segmenté, la numérotation des items découpés dans le texte de départ permet de constituer un système de coordonnées sur le texte dans lequel chaque item est repéré par son numéro d'ordre. Nous appelons ce système de coordonnées sur la séquence textuelle : la **Trame** textométrique. Ce même système de coordonnées permet de définir et de localiser, au sein du corpus, des zones textuelles (zones formées par une suite d'items consécutifs, entre la position x_1 la position x_2, réunion d'un certain nombre de zones de ce type, etc.).*

*La définition d'une **Trame** textométrique sur un corpus de textes permet de décrire les systèmes de zones qui correspondent aux contenants de l'analyse textométrique (parties, paragraphes, phrases, sections, chapitres etc.). On peut rassembler les descriptions relatives aux systèmes de contenants dans une structure de données particulière le **Cadre** textométrique.*



*Nous appelons **Trameur** l'outil informatique qui permet de construire une ressource textométrique **Trame**/**Cadre**, à partir d'un ensemble de textes rassemblés en corpus, selon les principes définis plus haut. La partie **Trame** de la ressource textuelle produite par le trameur est constituée par la suite des items isolés lors de l'opération de segmentation. La partie **Cadre** rassemble les données relatives aux différents découpages réalisés sur le corpus ainsi que les annotations déjà projetées sur les contenants ou les contenus constitués lors des étapes précédentes. La transmission d'une ressource textuelle constituée sous la forme **Trame**/**Cadre** constitue une solution suffisante pour servir de base à toute exploration textométrique ultérieure.*

Source : [Söze-Duval, 2008], Keyser Söze-Duval. *Pour une textométrie opérationnelle*. (DOC)

3 Présentation générale

Le ***Trameur*** : Programme de génération puis de gestion de la ***Trame*** et du ***Cadre*** d'un texte (i.e découpage en unité et partitionnement du texte : le métier textométrique) pour des opérations textométriques (ventilation des unités, carte des sections, cooccurrence, spécificité, AFC...).

Le ***Trameur*** intègre le programme *treetagger*[1] : système d'étiquetage automatique des catégories grammaticales des mots avec lemmatisation.

Il permet aussi de génèner et de gérer des annotations multiples sur les unités du texte (et de traiter les niveaux d'annotations visés).

Ce document présente tout d'abord la « *version Tk* » du ***Trameur*** puis la « *version console* ».

TOUTES LES DISTRIBUTIONS "COMPLETES" DU TRAMEUR INTEGRENT PAR DEFAUT LES MODULES COMPLEMENTAIRES : *R*, *TREETAGGER* ET *PAJEK*

TOUTES LES DISTRIBUTIONS "COMPLETES" INTEGRENT UNE RESSOURCE D'ANNOTATION POUR TREETAGGER ADAPTEE AU FRANÇAIS ORAL, CELLE-CI A ETE REALISEE PAR CHRISTOPHE BENZITOUN (ATILF) (source : http://cnrtl.fr/corpus/perceo/)

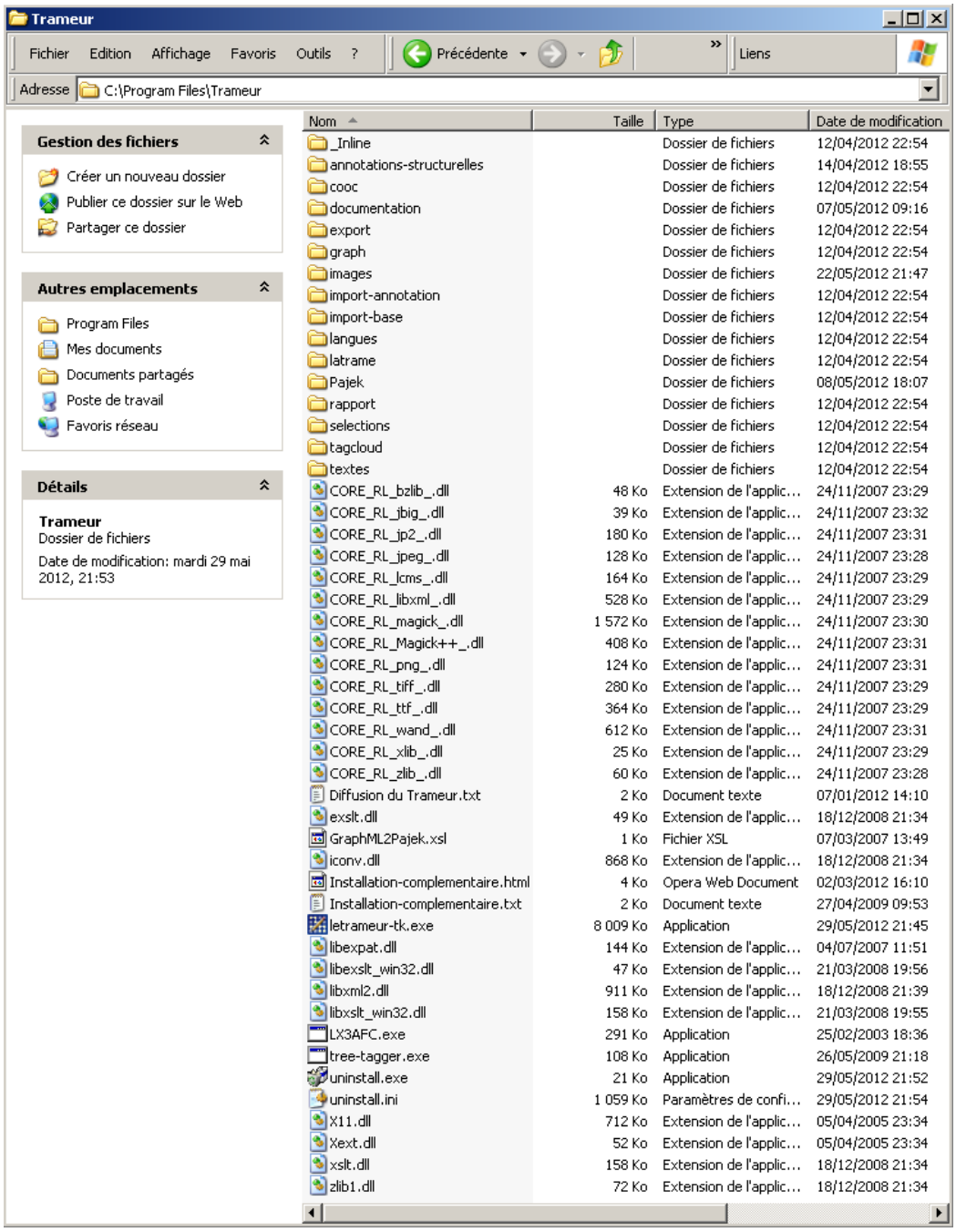
TOUTES LES DISTRIBUTIONS "ALLEGEES" N'INTEGRENT PAS LES MODULES COMPLEMENTAIRES : *TREETAGGER* ET *PAJEK*. POUR LES INSTALLER, SUIVRE LE MODE D'EMPLOI DISPONIBLE EN LIGNE.

[1] http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html

4 Installation

4 .1 Installation du Trameur

Après installation du programme `setup-trameur.exe`, l'arborescence de fichiers de l'application a l'allure suivante :


Trameur
Fichier Edition Affichage Favoris Outils ?
Précédente
Liens
Adresse C:\Program Files\Trameur
Gestion des fichiers
Créer un nouveau dossier
Publier ce dossier sur le Web
Partager ce dossier
Autres emplacements
Program Files
Mes documents
Documents partagés
Poste de travail
Favoris réseau
Détails
Trameur
Dossier de fichiers
Date de modification: mardi 29 mai 2012, 21:53
Nom Taille Type Date de modification
_Inline Dossier de fichiers 12/04/2012 22:54
annotations-structurelles Dossier de fichiers 14/04/2012 18:55
cooc Dossier de fichiers 12/04/2012 22:54
documentation Dossier de fichiers 07/05/2012 09:16
export Dossier de fichiers 12/04/2012 22:54
graph Dossier de fichiers 12/04/2012 22:54
images Dossier de fichiers 22/05/2012 21:47
import-annotation Dossier de fichiers 12/04/2012 22:54
import-base Dossier de fichiers 12/04/2012 22:54
langues Dossier de fichiers 12/04/2012 22:54
latrame Dossier de fichiers 12/04/2012 22:54
Pajek Dossier de fichiers 08/05/2012 18:07
rapport Dossier de fichiers 12/04/2012 22:54
selections Dossier de fichiers 12/04/2012 22:54
tagcloud Dossier de fichiers 12/04/2012 22:54
textes Dossier de fichiers 12/04/2012 22:54
CORE_RL_bzlib_.dll 48 Ko Extension de l'applic... 24/11/2007 23:29
CORE_RL_jbig_.dll 39 Ko Extension de l'applic... 24/11/2007 23:32
CORE_RL_jp2_.dll 180 Ko Extension de l'applic... 24/11/2007 23:31
CORE_RL_jpeg_.dll 128 Ko Extension de l'applic... 24/11/2007 23:28
CORE_RL_lcms_.dll 164 Ko Extension de l'applic... 24/11/2007 23:29
CORE_RL_libxml_.dll 528 Ko Extension de l'applic... 24/11/2007 23:29
CORE_RL_magick_.dll 1 572 Ko Extension de l'applic... 24/11/2007 23:30
CORE_RL_Magick++_.dll 408 Ko Extension de l'applic... 24/11/2007 23:31
CORE_RL_png_.dll 124 Ko Extension de l'applic... 24/11/2007 23:31
CORE_RL_tiff_.dll 280 Ko Extension de l'applic... 24/11/2007 23:29
CORE_RL_ttf_.dll 364 Ko Extension de l'applic... 24/11/2007 23:29
CORE_RL_wand_.dll 612 Ko Extension de l'applic... 24/11/2007 23:31
CORE_RL_xlib_.dll 25 Ko Extension de l'applic... 24/11/2007 23:29
CORE_RL_zlib_.dll 60 Ko Extension de l'applic... 24/11/2007 23:28
Diffusion du Trameur.txt 2 Ko Document texte 07/01/2012 14:10
exslt.dll 49 Ko Extension de l'applic... 18/12/2008 21:34
GraphML2Pajek.xsl 1 Ko Fichier XSL 07/03/2007 13:49
iconv.dll 868 Ko Extension de l'applic... 18/12/2008 21:34
Installation-complementaire.html 4 Ko Opera Web Document 02/03/2012 16:10
Installation-complementaire.txt 2 Ko Document texte 27/04/2009 09:53
letrameur-tk.exe 8 009 Ko Application 29/05/2012 21:45
libexpat.dll 144 Ko Extension de l'applic... 04/07/2007 11:51
libexslt_win32.dll 47 Ko Extension de l'applic... 21/03/2008 19:56
libxml2.dll 911 Ko Extension de l'applic... 18/12/2008 21:39
libxslt_win32.dll 158 Ko Extension de l'applic... 21/03/2008 19:55
LX3AFC.exe 291 Ko Application 25/02/2003 18:36
tree-tagger.exe 108 Ko Application 26/05/2009 21:18
uninstall.exe 21 Ko Application 29/05/2012 21:52
uninstall.ini 1 059 Ko Paramètres de confi... 29/05/2012 21:54
X11.dll 712 Ko Extension de l'applic... 05/04/2005 23:34
Xext.dll 52 Ko Extension de l'applic... 05/04/2005 23:34
xslt.dll 158 Ko Extension de l'applic... 18/12/2008 21:34
zlib1.dll 72 Ko Extension de l'applic... 18/12/2008 21:34

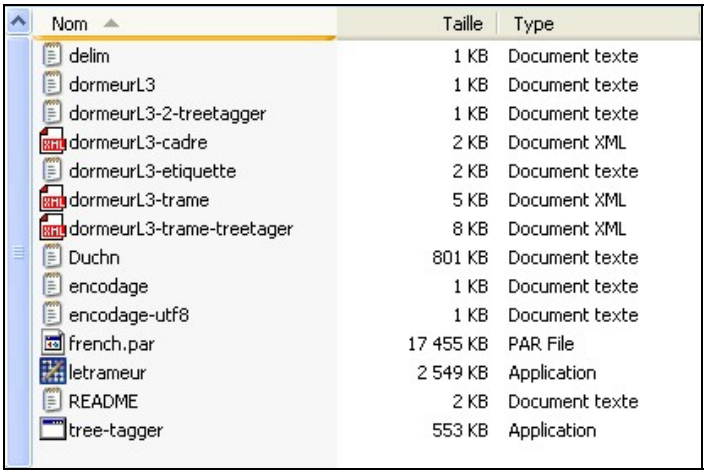
Descriptif des principales ressources du dossier d'installation :

- Le dossier `_Inline` contient des bibliothèques externes pour le ***Trameur***
- Le dossier `annotations-structuelles` contient des ressources illustrant la construction de Graphe d'annotation (cf infra)
- Le dossier `cooc` contient des ressources pour le calcul des cooccurrents (stop-liste)

- Le dossier `documentation` contient ce même document en version PDF et en version HTML
- Le dossier `export`[2] est utilisé pour stocker les exports de fichier générés par le ***Trameur***
- Le dossier `graph` est utilisé pour stocker les graphiques générés par le ***Trameur***
- Le dossier `images` contient les différentes icônes utilisées par le ***Trameur***
- Les dossiers `import` contiennent des modèles de fichier d'importation de bases textométriques ou de système d'annotations sur les fichiers de travail utilisés dans cette documentation.
- Le dossier `langues` contient différents fichiers de paramétrage de langue nécessaire pour *treetagger*
- Le dossier `Pajek` contient le programme du même nom et les ressources associées
- Le dossier `rapport` contient les éléments utilisés pour générer les rapports produit par le ***Trameur***
- Le dossier `textes` contient des exemples de fichier de travail
- Le programme *tree-tagger.exe* est lancé par le ***Trameur*** pour étiqueter le texte de travail
- Le fichier `GraphML2Pajek.xsl` est utilisé par le ***Trameur*** pour exporter une liste de séquences de termes associées à un patron au format *Pajek*
- Le programme le ***Trameur*** enfin : `letrameur-tk.exe`. Pour le lancer : double-clic sur son icône

4 .2 Installation du Trameur (version console)

Après avoir décompressé l'archive contenant le programme, l'arborescence de fichiers de l'application a l'allure suivante :


Nom
Taille
Type
delim 1 KB Document texte
dormeurL3 1 KB Document texte
dormeurL3-2-treetagger 1 KB Document texte
dormeurL3-cadre 2 KB Document XML
dormeurL3-etiquette 2 KB Document texte
dormeurL3-trame 5 KB Document XML
dormeurL3-trame-treetager 8 KB Document XML
Duchn 801 KB Document texte
encodage 1 KB Document texte
encodage-utf8 1 KB Document texte
french.par 17 455 KB PAR File
letrameur 2 549 KB Application
README 2 KB Document texte
tree-tagger 553 KB Application

Le programme le ***Trameur*** se nomme ici : `letrameur.exe`. Pour le lancer, démarrer une fenêtre de commandes et suivre le mode d'emploi contenu dans le fichier README.txt ou dans ce document.

[2] Un dossier de même nom sera créé dans le dossier contenant la base en cours de traitement. Des fichiers d'export y seront stockés (idem pour les dossiers `graph`, `logR`, `rapport` : ces dossiers stockeront des données générées au cours de l'analyse de la base)

5 Interface du *Trameur*

L'interface du programme est composée d'une fenêtre graphique disposant de différents onglets, chacun permettant d'activer différentes fonctionnalités.

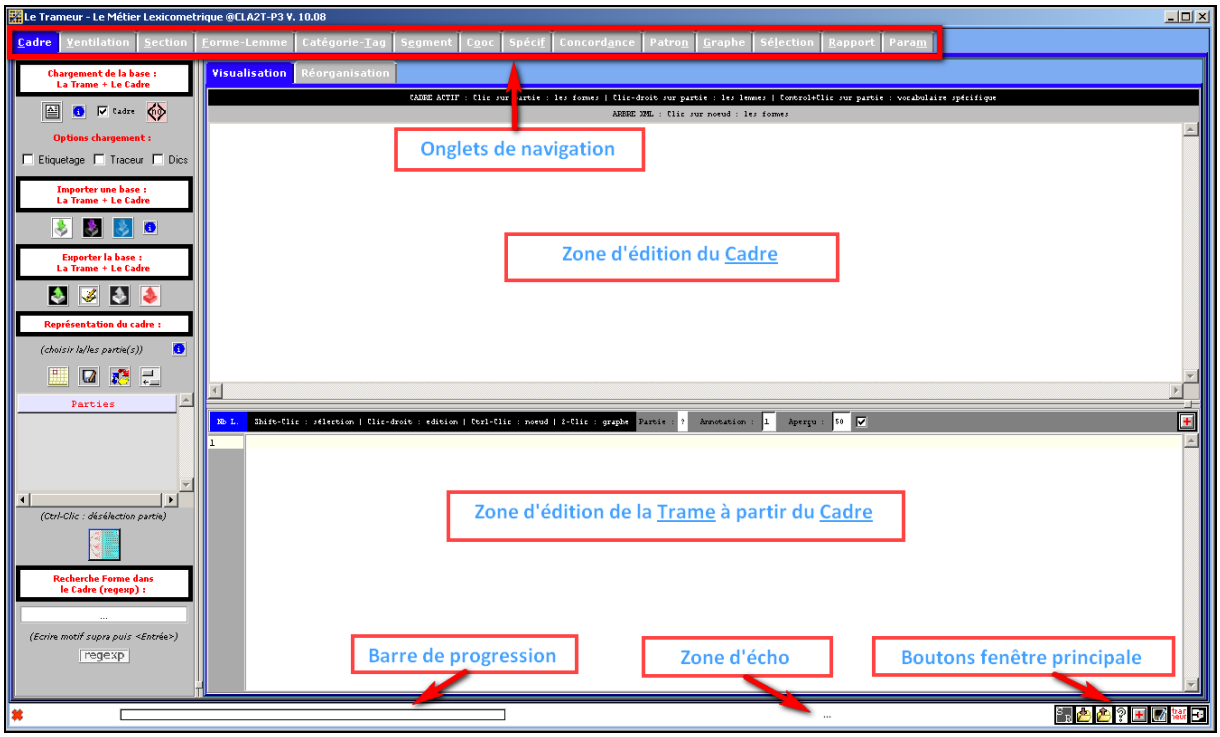


Figure 1 : *Le Trameur* Fenêtre principale

L'onglet **CADRE** visible au démarrage permet de charger la base de travail.
L'onglet **VENTILATION** donne accès aux fonctionnalités liées à la génération ou à l'affichage des graphes de ventilation.
L'onglet **SECTION** donne accès aux fonctionnalités liées à la *Carte des Sections*.
L'onglet **FORME-LEMME** donne accès aux dictionnaires des formes et des lemmes et aux fonctionnalités associées.
L'onglet **CATEGORIE-TAG** donne accès aux dictionnaires des catégories et des annotations complémentaires et aux fonctionnalités associées.
L'onglet **SEGMENT** donne accès aux fonctionnalités liées au calcul des segments répétés.
L'onglet **COOC** donne accès aux fonctionnalités liées aux calculs des cooccurrents et polycooccurrents.
L'onglet **STAT** donne accès aux fonctionnalités liées aux calculs statistiques (spécificités, AFC etc.).
L'onglet **CONCORDANCE** donne accès aux fonctionnalités liées aux concordances.
L'onglet **PATRON** donne accès aux fonctionnalités liées aux calculs d'extraction de patron.
L'onglet **GRAPHE** donne accès aux fonctionnalités liées aux graphes de relation entre les items de la ***Trame***.
L'onglet **RELATION** (non visible dans la figure 1) donne accès aux fonctionnalités liées aux relations de dépendance entre les items de la ***Trame***.
L'onglet **SELECTION** donne accès aux fonctionnalités du *Gestionnaire de Sélection*.
L'onglet **RAPPORT** donne accès aux fonctionnalités liées à la génération du rapport d'exploration d'une base.
L'onglet **PARAM** donne accès aux fonctionnalités permettant de paramétrer le chargement d'une base de travail.

6 Descriptif des icônes[3]

Icône	Fonction	Localisation
	Chargement de la base	Onglet **CADRE**
	Chargement de la représentation graphique du ***Cadre***	Onglet **CADRE**
	Chargement de la représentation graphique de la Carte des sections	Onglet **SECTION**
	Suppression de la représentation graphique du ***Cadre***	Onglet **CADRE**
	Sauvegarde de la représentation graphique du ***Cadre***	Onglet **CADRE**
	Chargement de la représentation arborescente d'un document au format XML	Onglet **CADRE**
	Ajouter un élément au rapport	Fenêtre principale, tous les onglets
	Enregistrer le rapport	Fenêtre principale, onglet **RAPPORT**
	Export de la base textométrique (***Trame***+***Cadre***)	Onglet **CADRE**
	Export (***Trame***+***Cadre***) : format TXT et XML	Onglet **CADRE**
	Import d'une base textométrique (***Trame***+***Cadre***)	Onglet **CADRE**
	Export du fichier ***Cadre*** et du fichier ***Trame***	Fenêtre principale
	Importation d'une annotation sur une ***Trame***	Fenêtre principale
	Export de la concordance au format HTML	Onglet **CONCORDANCE**
	Documentation du programme au format HTML	Fenêtre principale
	Calcul des segments répétés	Fenêtre principale, onglet **SEGMENT**
	Génération de liste de séquences de termes associées à un patron	Onglet **PATRON**
	Export de liste des séquences de termes associées à un patron au format *Pajek*	Onglet **PATRON**
	Génération de la courbe d'accroissement du vocabulaire	Onglet **VENTILATION**
	Calcul des spécificités (cooc) dans les sections contenant le motif	Onglet **SECTION**
	Calcul des spécificités (cooc) dans les sections ne contenant pas le motif	Onglet **SECTION**
	Calcul des spécificités dans les sections sélectionnées	Onglet **SECTION**
	Calcul des mots spécifiques d'une série sections sélectionnées	Onglet **SECTION**
	Edition des items	Onglets **FORME-LEMME, CATEGORIE-TAG**
	Ventilation des items	Onglets **FORME-LEMME, CATEGORIE-TAG, SEGMENT**
	Concordance des items	Onglets **FORME-LEMME, CATEGORIE-TAG, SEGMENT**
	Projection sur le ***Cadre*** des items	Onglets **FORME-LEMME, CATEGORIE-TAG, SEGMENT**
	Projection sur la carte des sections des items	Onglets **FORME-LEMME, CATEGORIE-TAG, SEGMENT, RELATION**
	Calcul des cooccurrents et des poly-cooccurrents	Onglet **COOCS**
	Calcul du tableau lexical d'une partition	Onglet **STAT**
	Calcul AFC	Onglet **STAT**
	Calcul des Principales Caractéristiques Lexicométriques d'une partition	Onglet **STAT**

[3] Le tableau ci-dessous intègre une grande partie des boutons disponibles dans l'interface du logiciel.

7 Chargement de la base : mise au jour du Cadre et de la Trame

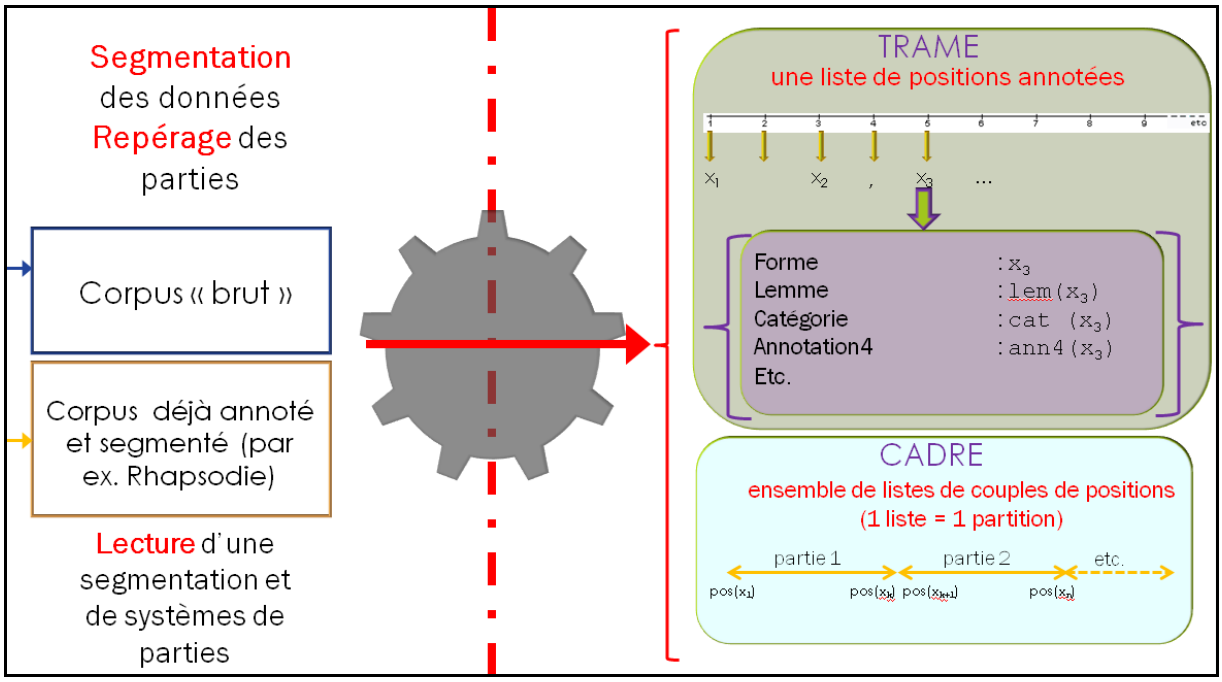
> *A partir d'un texte segmenté, la numérotation des items découpés dans le texte de départ permet de constituer un système de coordonnées sur le texte dans lequel chaque item est repéré par son numéro d'ordre. Nous appelons ce système de coordonnées sur la séquence textuelle : la* ***Trame*** *textométrique. Ce même système de coordonnées permet de définir et de localiser, au sein du corpus, des zones textuelles (zones formées par une suite d'items consécutifs, entre la position x_1 la position x_2, réunion d'un certain nombre de zones de ce type, etc.). La définition d'une* ***Trame*** *textométrique sur un corpus de textes permet de décrire les systèmes de zones qui correspondent aux contenants de l'analyse textométrique (parties, paragraphes, phrases, sections, chapitres etc.). On peut rassembler les descriptions relatives aux systèmes de contenants dans une structure de données particulière le* ***Cadre*** *textométrique. (Söze-Duval 2008)*

Une base textométrique peut être chargée dans Le ***Trameur*** :

1. **en créant une nouvelle base** (à partir de fichiers dont le format est défini *infra*)
2. **ou en important une base déjà créée** (*cf* export d'une base textométrique)

« **Créer une nouvelle base** » déclenche la sélection du fichier visé puis son chargement (avec ou sans étiquetage via *Treetagger*) *i.e* génération d'une segmentation et du repérage des parties (*cf infra* format de fichier).

« **Importer une base** » permet de charger une base textométrique déjà constituée (par exemple une base déjà exportée par le logiciel) *i.e* une base intégrant la description d'une ***Trame*** et d'un ***Cadre***.



7.1 Chargement d'une nouvelle base

7.1.1 Paramétrage

L'onglet **PARAM** contient différents champs permettant de paramétrer le chargement de la base :

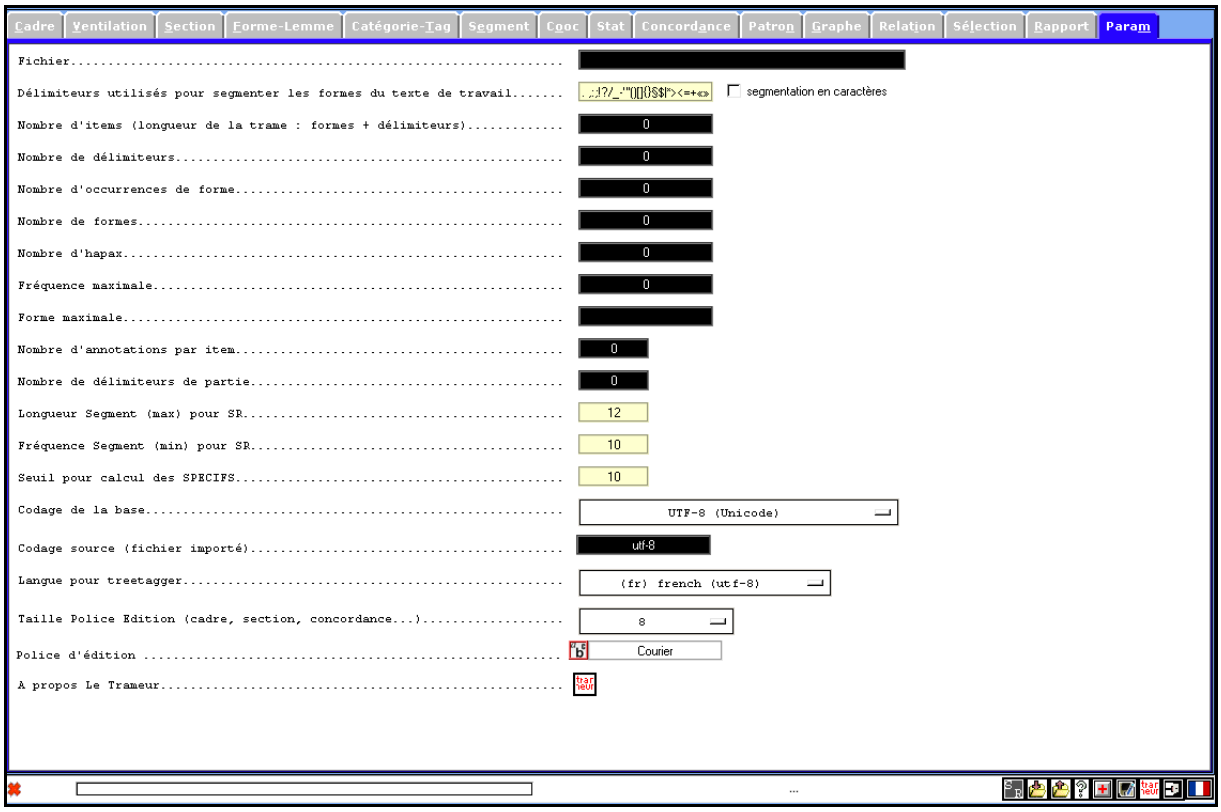


Figure 2 : *Le Trameur* Onglet PARAM

On peut y définir les délimiteurs utilisés pour segmenter le texte, l'encodage du texte à traiter, le fichier de langue à utiliser pour l'étiquetage via *treetagger* et les paramètres pour le calcul des segments répétés. On peut aussi y modifier la police d'édition du texte chargé (nécessaire pour le traitement de certaines langues). Les autres champs sont des indicateurs renseignés par le programme à l'issue du chargement de la base.

7.1.2 Chargement

L'onglet **CADRE** contient :

- une zone complémentaire pour paramétrer le chargement de la base
- une fenêtre d'édition du ***Cadre*** de la base (représentation graphique de l'enchâssement des parties)
- une fenêtre d'édition d'une partie sélectionnée dans le ***Cadre***

2 cases à cocher permettent de déclencher :

a. l'étiquetage du texte traité (avec *treetagger*),

b. l'impression des différents dictionnaires (*Dics*) : L'activation de *Dics* conduit à l'impression des différents dictionnaires disponibles à l'issue du chargement.

Pour charger une *nouvelle base* : activez le bouton

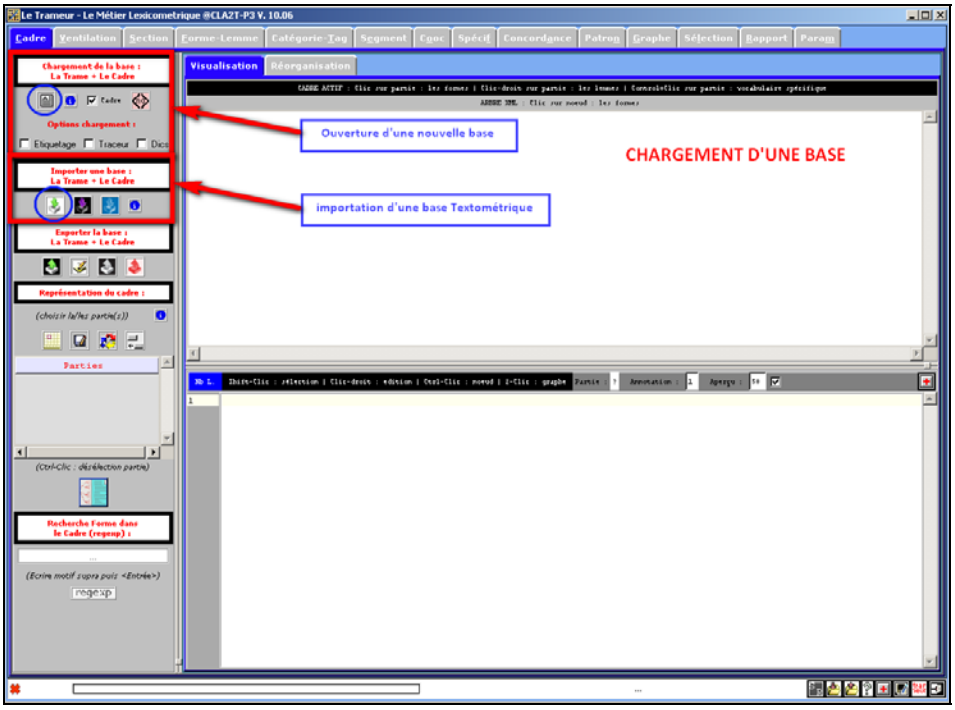


Figure 3 : *Le Trameur* Onglet CADRE

7.1.3 Etiquetage du texte

Le ***Trameur*** intégre (dans sa version complète) le programme *treetagger*[4] : un outil qui permet d'annoter un texte avec des informations sur les parties du discours (genre de mots: noms, verbes, infinitifs et particules) et des informations de lemmatisation. Pour activer l'étiquetage du texte chargé au démarrage, il est nécessaire, avant de charger le texte, de cocher la case à cocher « étiquetage » (*cf* onglet **CADRE**) et de choisir le fichier paramètre de langue à utiliser pour réaliser cet étiquetage (*cf* onglet **PARAM**) :

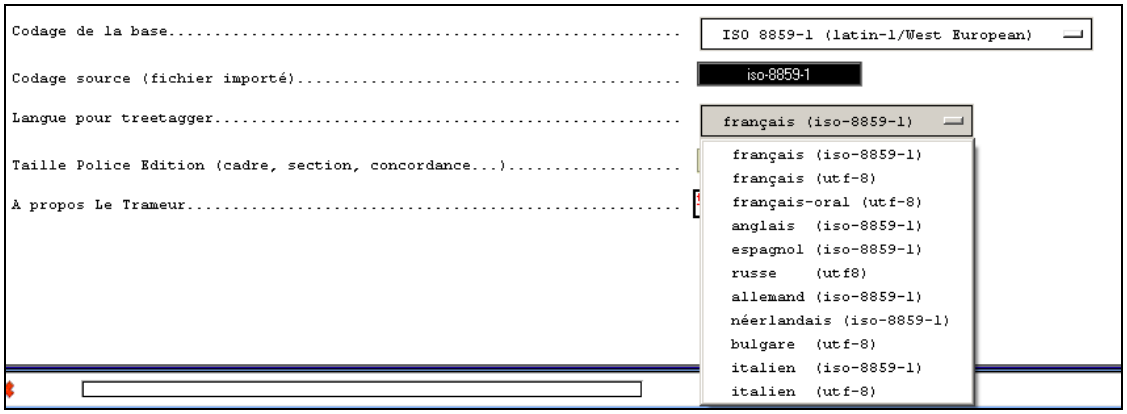


Figure 4 : *Le Trameur* Onglet PARAM : choix de langue pour Treetagger

Les langues utilisables sont visibles dans la figure précédente. La plupart des fichiers paramètres de langue associés ont été récupérés sur le site de *treetagger*. Le fichier « français oral » est issu du projet PERCEO[5] (ATILF (CNRS - Université de Lorraine) & INIST (CNRS) – LIPN), responsable scientifique Christophe Benzitoun (ATILF).

[4] http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/
[5] http://cnrtl.fr/corpus/perceo/

7.1.4 Inventaire typographique

Après avoir chargé une nouvelle base, il est possible de visualiser l'ensemble des caractères composant le fichier de travail (module *Inventaire Typographique*), on peut éventuellement modifier certains caractères ou modifier le statut de certains caractères (délimiteur). Pour illustrer ces fonctionnalités, on considère que la nouvelle base chargée a été construite à partir du fichier suivant :

```
<STRUCTURE="TITRE">Le dormeur du val § £
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1">C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2">Accrochant follement aux herbes des haillons
<LIGNE="VERS3">D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4">Luit : c' est un petit val qui mousse de rayons. §
<LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
<LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
<LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
<LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut. §
<LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
<LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
<LIGNE="VERS11">Nature, berce-le chaudement : il a froid. §
<LIGNE="VERS12">Les parfums ne font pas frissonner sa narine ;
<LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
<LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit. § £
<STRUCTURE="AUTEUR">Arthur Rimbaud § £
```

Une fois la base chargée, l'*inventaire typographique* produit le résultat suivant :

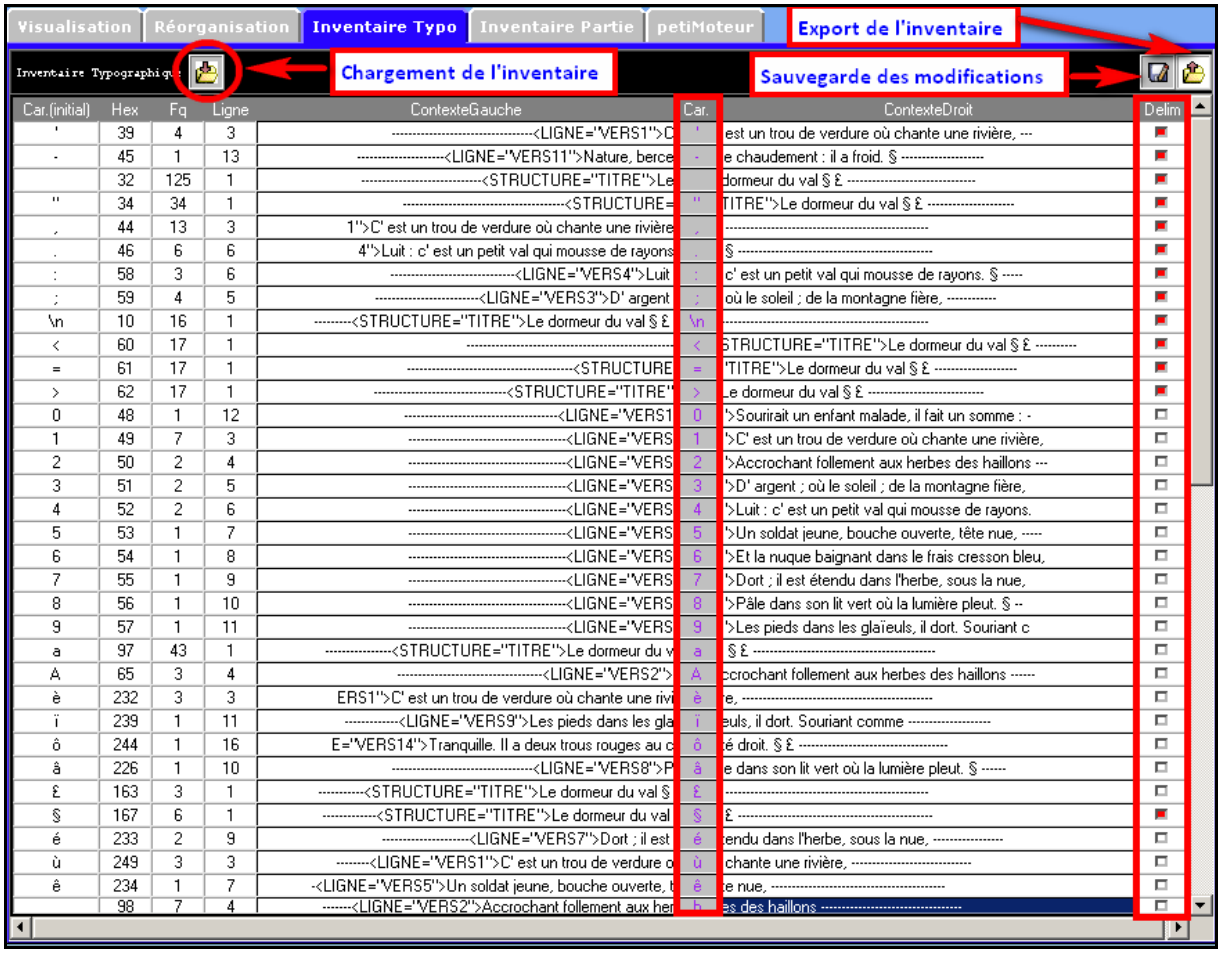


Figure 5a : *Le Trameur* Inventaire Typographique

La **sixième colonne** met au jour les caractères en contexte (contextes gauche et droit) dans une fenêtre d'une cinquantaine de caractères : chaque cellule de cette colonne est éditable (clic sur la cellule).

La **huitième colonne** indique le statut du caractère : si la case est rouge, le caractère est un délimiteur. Un clic sur la case permet de changer le statut du caractère visé.

Les modifications étant réalisées, la sauvegarde du fichier (en tenant compte ses modifications réalisées) déclenche automatiquement le rechargement de la base dans le ***Trameur*** et la mise à jour éventuelle de la liste des délimiteurs (*cf* onglet **PARAM**).

Remarque : le processus de correction modifie le fichier initial, une sauvegarde de ce fichier est réalisée au préalable dans le même dossier.

Pour illustrer 2 fonctionnalités, on utilise *infra* un corpus issu du module *petiMoteur* disponible dans l'onglet **CADRE**.

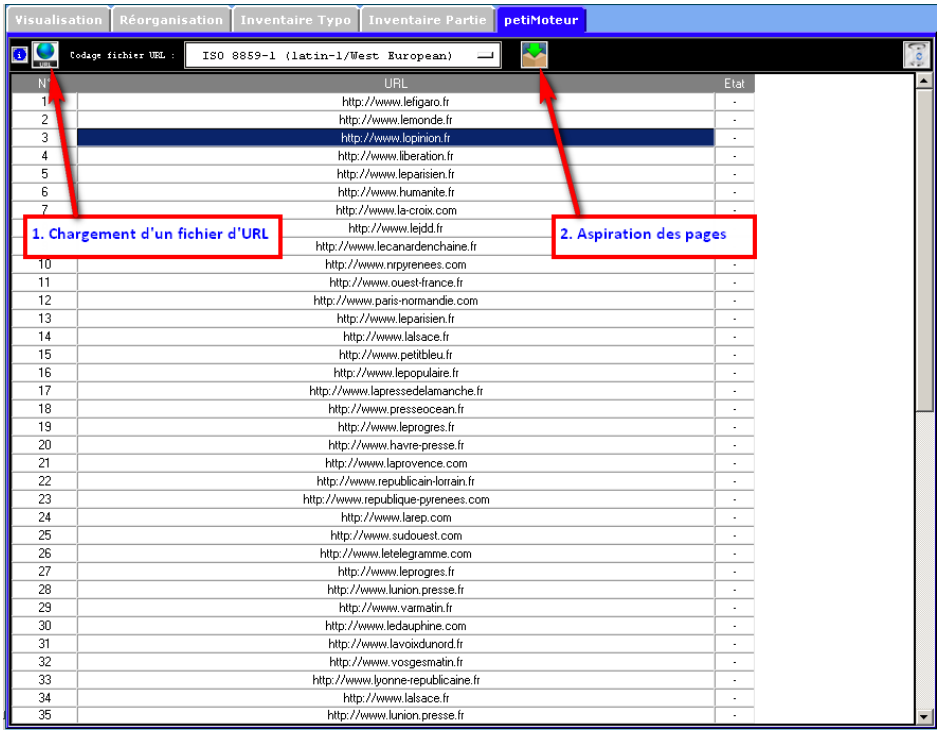


Figure 5b : *Le Trameur* Inventaire Typographiquesur petitMoteur (1)

Ce module permet de réaliser un corpus à partir d'une liste d'URLs (ici une liste de 60 urls pointant vers des pages d'accueil de journaux français en ligne). Une fois le fichier d'URLs chargé (1), le module d'aspiration (2) construit un fichier concaténant les contenus textuels des différentes URLs. Le fichier produit (text_URLS.txt dans le même dossier que le fichier des URLs) peut ensuite être chargé dans le ***Trameur***.

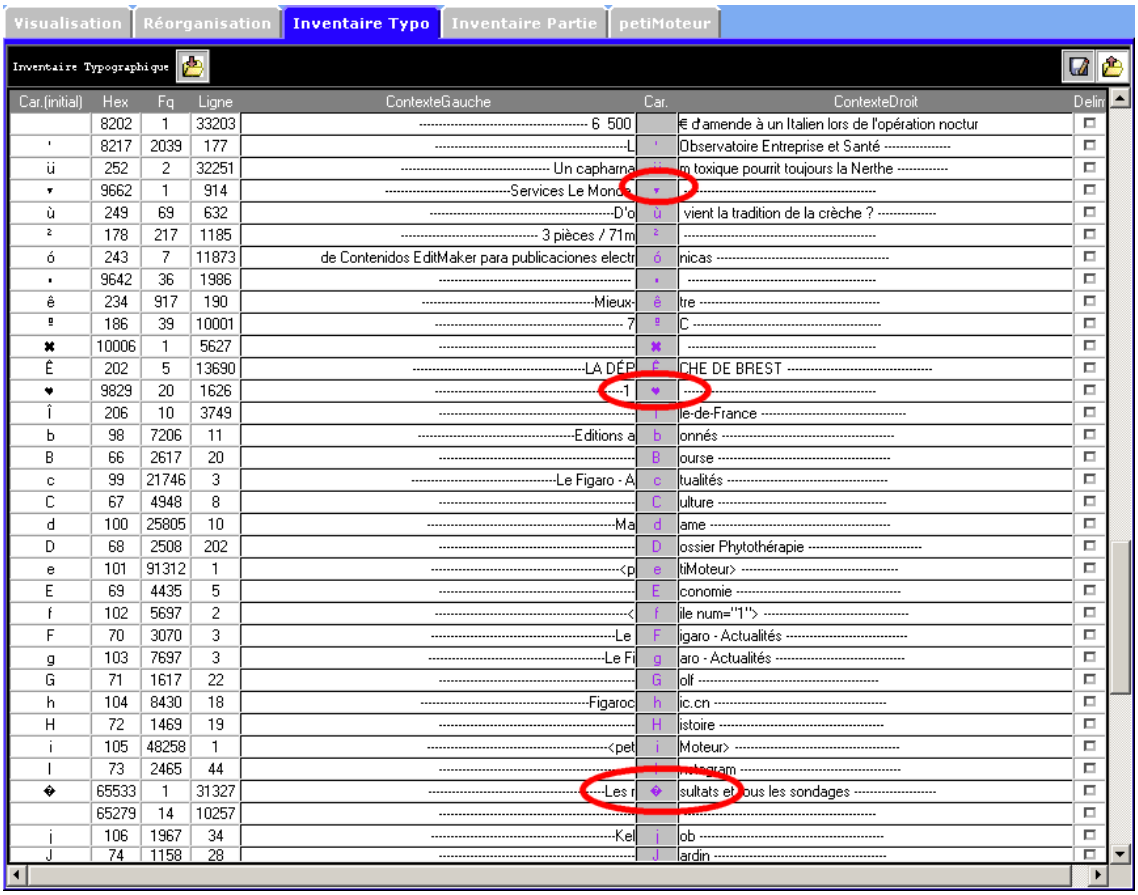


Figure 5c : *Le Trameur* Inventaire Typographiquesur petitMoteur (2)

L'inventaire révèle ici de nombreux caractères sur lesquels des modifications sont envisageables : erreurs de codage par exemple, attribution de statut de délimiteur etc.

7.1.5 Inventaire des parties

De la même manière, il est possible de construire un inventaire de toutes les balises utilisées pour définir le partitionnement du fichier chargé. Une fois la base chargée, l'*inventaire des parties* a l'allure suivante :

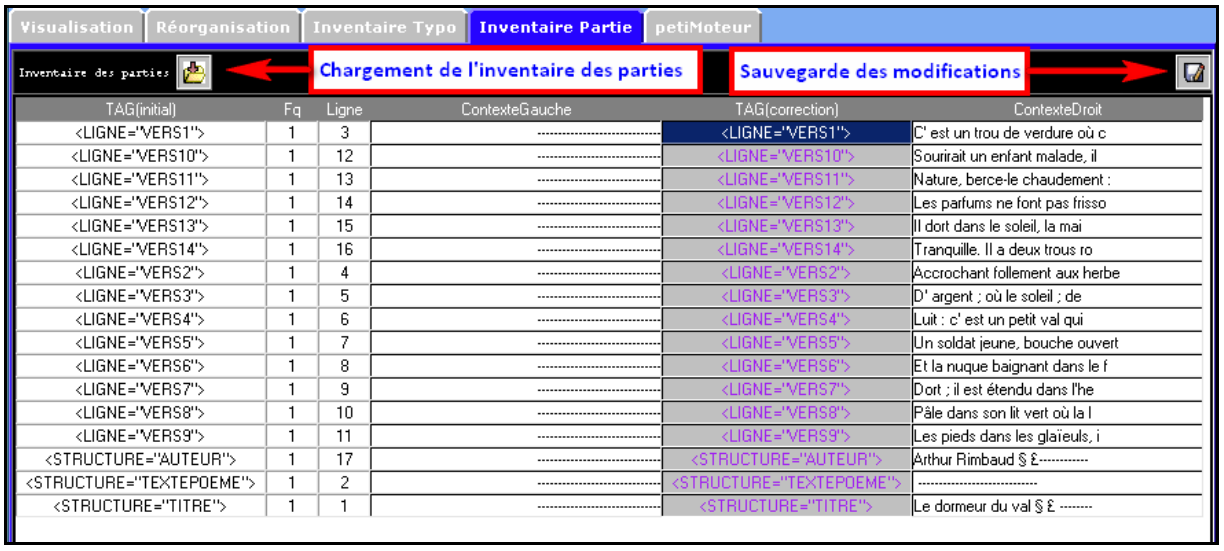


Figure 5d : *Le Trameur* Inventaire des parties

La **cinquième colonne** met au jour chaque balise en contexte (contextes gauche et droit) dans une fenêtre d'une cinquantaine de caractères : chaque cellule de cette colonne est éditable (clic sur la cellule).

Les modifications étant réalisées, la sauvegarde du fichier (en tenant compte ses modifications réalisées) déclenche automatiquement le rechargement de la base dans le ***Trameur***.

Remarque : le processus de correction modifie le fichier initial, une sauvegarde de ce fichier est réalisée au préalable dans le même dossier.

Ce processus peut être utile pour corriger le fichier chargé en supprimant notamment les caractères < et > résiduels : *cf* ci-dessous. Ce module permet aussi de renommer les balises ; ci-dessous, 2 opérations de renommage sont réalisées, l'une d'entre elles s'accompagne aussi d'un ajout d'une nouvelle balise.

Ci-dessous, un corpus issu du programme *Gromoteur*[6] (même liste d'URLs que précédemment) dont l'*inventaire des parties* produit ceci :

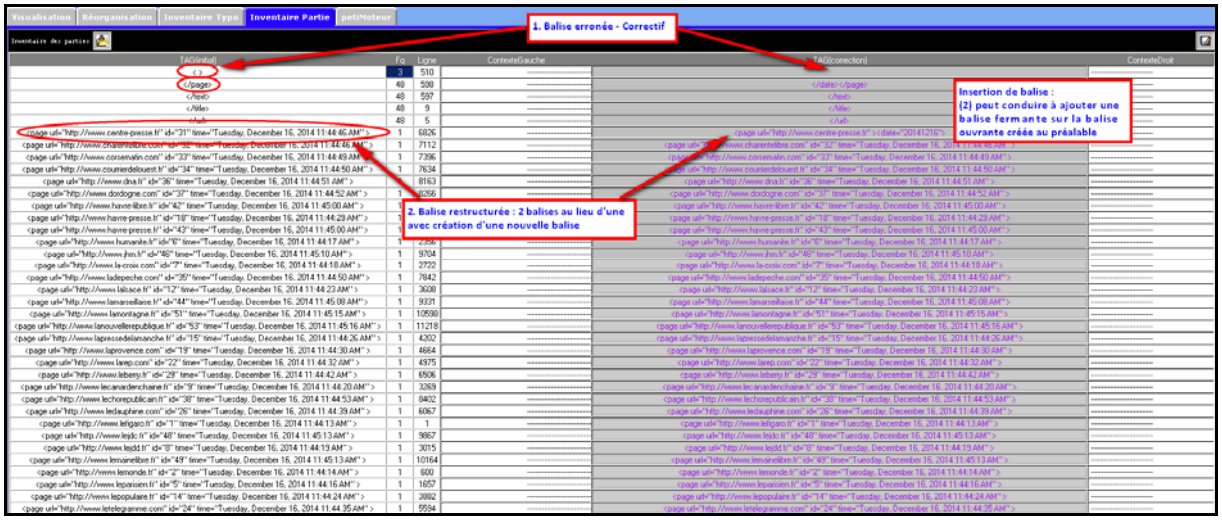

Visualisation
Réorganisation
Inventaire Typo
Inventaire Partie
petiMoteur
1. Balise erronée - Correctif
2. Balise restructurée : 2 balises au lieu d'une avec création d'une nouvelle balise
Insertion de balise :
(2) peut conduire à ajouter une balise fermante sur la balise ouvrante créée au préalable

Figure 5e : *Le Trameur* Inventaire des parties sur Gromoteur

L'inventaire peut conduire à supprimer des balises « erronées » ou à modifier les balises existantes pour redéfinir le partitionnement de la base après sauvegarde des modifications.

[6] http://gromoteur.ilpga.fr/

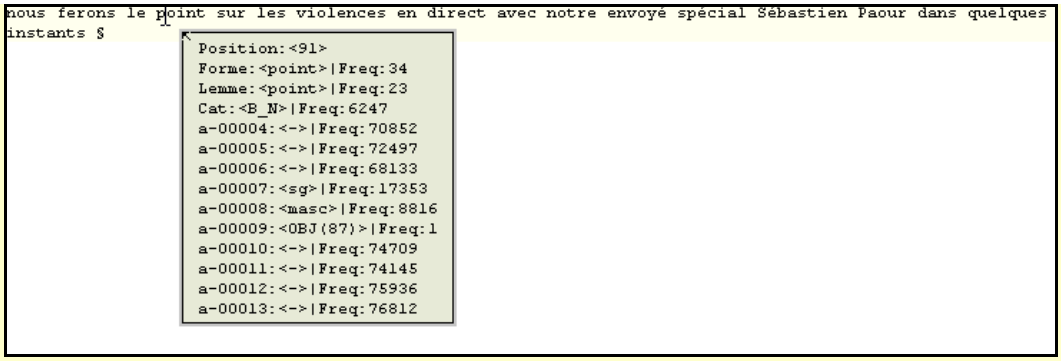
7.2 Importer une base

On peut aussi charger une base en important une base textométrique existante (une ***Trame*** et un ***Cadre***). **On active cette fonctionnalité *via* le bouton idoine**

Le format du fichier d'importation d'une base est présenté dans la section « Exporter une base textométrique ».

On trouvera des exemples de bases textométriques déjà constituées sur la page web du projet. Ci-dessous, quelques bases utilisées dans ce document :

- **Base "Rhapsodie2Trameur"** construite à partir des ressources développées dans le cadre du projet **Rhapsodie**. Chaque item de la *Trame* est associé à 13 niveaux d'annotation (micro-syntaxe).



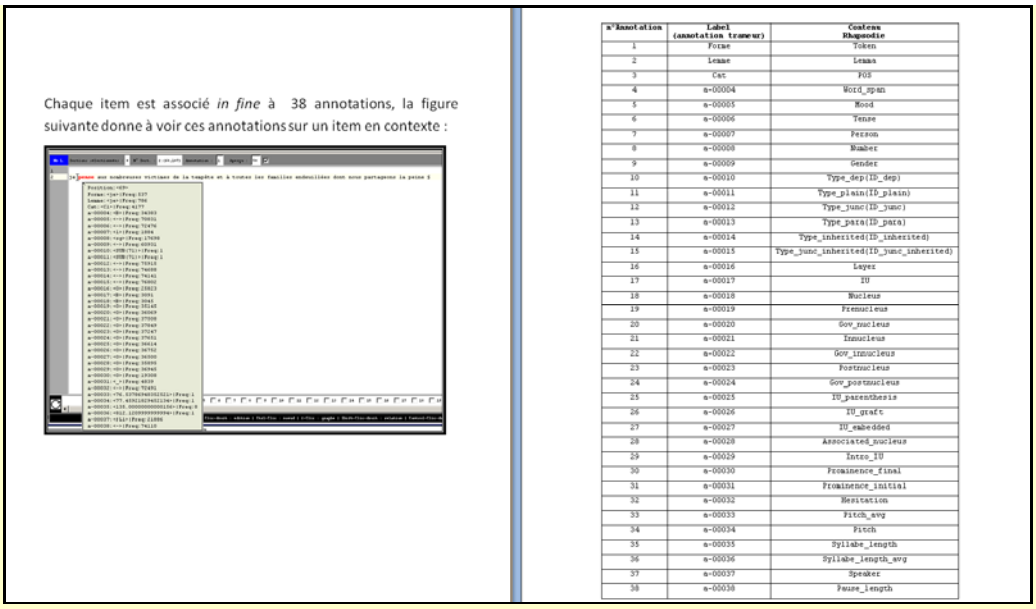
SOURCES : projet Rhapsodie

Fichiers de codage microsyntaxique version béta 10/13 (zip)

Descriptif et sources des annotations :
Annotations Rhapsodie pour le Trameur (pdf) : présentation du processus de transcodage des annotations Rhapsodie pour construire une base Textométrique ; présentation des différents processus de traitements des annotations de dépendance.

Open publication - Free publishing
Open publication - Free publishing

- **Base "Rhapsodie2Trameur" (v4) (màj : 19.07.2014)** construite à partir des ressources développées dans le cadre du projet **Rhapsodie**. Chaque item de la *Trame* est associé à 38 niveaux d'annotation (prosodie, micro et macro-syntaxe).



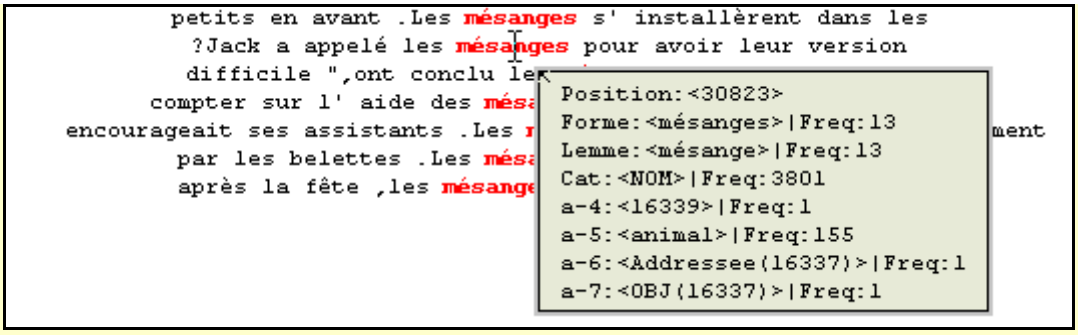
n°Annotation	Label (annotation trameur)	Contenu Rhapsodie
1	Forme	Token
2	Lemme	Lemma
3	Cat	POS
4	a-00004	Word_span
5	a-00005	Mood
6	a-00006	Tense
7	a-00007	Person
8	a-00008	Number
9	a-00009	Gender
10	a-00010	Type_dep(ID_dep)
11	a-00011	Type_plain(ID_plain)
12	a-00012	Type_junc(ID_junc)
13	a-00013	Type_para(ID_para)
14	a-00014	Type_inherited(ID_inherited)
15	a-00015	Type_junc_inherited(ID_junc_inherited)
16	a-00016	Layer
17	a-00017	IU
18	a-00018	Nucleus
19	a-00019	Prenucleus
20	a-00020	Gov_nucleus
21	a-00021	Innucleus
22	a-00022	Gov_innucleus
23	a-00023	Postnucleus
24	a-00024	Gov_postnucleus
25	a-00025	IU_parenthesis
26	a-00026	IU_graft
27	a-00027	IU_embedded
28	a-00028	Associated_nucleus
29	a-00029	Intro_IU
30	a-00030	Prominence_final
31	a-00031	Prominence_initial
32	a-00032	Hesitation
33	a-00033	Pitch_avg
34	a-00034	Pitch
35	a-00035	Syllabe_length
36	a-00036	Syllabe_length_avg
37	a-00037	Speaker
38	a-00038	Pause_length

(clic pour vue détaillée)

SOURCES : projet Rhapsodie

Descriptif et sources des annotations :
Annotations Rhapsodie pour le Trameur (pdf) : présentation du processus de transcodage des annotations Rhapsodie pour construire une base Textométrique ; présentation des différents processus de traitements des annotations de dépendance.

- **Base "Corpus de contes français"** (Fairy Tales Corpus, FTC) construite à partir des ressources développées par Ismaïl El Maarouf. Chaque item de la *Trame* est associé à 7 niveaux d'annotation.



SOURCES : I. El Maarouf et J. Villaneau (2012). *A French Fairy Tale Corpus syntactically and semantically annotated*. In Proceedings of LREC 2012. Istanbul (Tk).
Pour les formes : copyright Momes.net, http://momes.net.
Lemme et POS via treetagger : fichier paramètre langue (français, utf-8) développé par Christophe Benzitoun (ATILF)
Pour les annotations : http://elmaarouf.legtux.org/Accueil.html

Descriptif et sources des annotations :
balise <f> (forme) : copyright Momes.net, http://elmaarouf.legtux.org/data/FTC_FORM.txt
balise <l> (lemme) : fichier paramètre langue (français oral, utf-8) développé par Christophe Benzitoun (ATILF) (source : http://cnrtl.fr/corpus/perceo/)
balise <c> (POS) : fichier paramètre langue (français oral, utf-8) développé par Christophe Benzitoun (ATILF)
Annotations complémentaires (http://elmaarouf.legtux.org/Accueil.html) :
- Identifiants de mots : http://elmaarouf.legtux.org/data/FTC_ID.txt (première balise <a>)
- Annotation référentielle : http://elmaarouf.legtux.org/data/FTC_REF.txt (seconde balise <a>)

- Annotation en rôles sémantiques pour 26 verbes : http://elmaarouf.legtux.org/data/FTC_SEMROLE.txt (troisième balise <a>)
- Annotation syntaxique en dépendance pour 137 verbes : http://elmaarouf.legtux.org/data/FTC_SYNROLE.txt (dernière balise <a>)

L'ensemble des annotations du corpus initial (+ lemme et POS) regroupées ici au format d'une base textométrique pour le **Trameur**.

8 Les formats d'entrée (nouvelle base)

Voir aussi : Fleury Serge, *Le Trameur. Propositions de description et d'implémentation des objets textométriques*, (PDF).

> Le texte précédent met au jour une description des objets textométriques et les méthodes mises en œuvre dans le ***Trameur*** pour travailler sur et avec ces objets dans une perspective textométrique. On y détaille aussi les opérations permises sur une *base textométrique* : format des données textuelles, modification dynamique de la *Trame*, correction ou ajout d'annotation etc.

8.1 Format texte brut

```
Le dormeur du val
C' est un trou de verdure où chante une rivière,
Accrochant follement aux herbes des haillons
D' argent ; où le soleil ; de la montagne fière,
Luit : c' est un petit val qui mousse de rayons.
```

Le texte contient une ***Trame*** mais pas de ***Cadre***. Le ***Trameur*** peut travailler sur la ***Trame*** du texte.

8.2 Format dit « Lexico »

Exemple n°1 :

```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
```

Le texte contient une ***Trame*** et un ***Cadre*** (balise de début explicite pour chaque partie, balise de fin implicite). Le ***Trameur*** peut travailler sur la ***Trame*** du texte et il est possible d'accéder au texte par son ***Cadre***.

8.3 Format XML

Exemple n°2 :

```
<?xml version="1.0" encoding="iso-8859-1"?>
<poeme>
<structure niveau="TITRE"> Le dormeur du val</structure>
<structure niveau="TEXTEPOEME">
<ligne type="VERS1"> C'est un trou de verdure où chante une rivière,</ligne>
<ligne type="VERS2"> Accrochant follement aux herbes des haillons</ligne>
<ligne type="VERS3"> D'argent ; où le soleil ; de la montagne fière,</ligne>
<ligne type="VERS4"> Luit : c'est un petit val qui mousse de rayons.</ligne>
</structure>
</poeme>
```

Exemple n°3 :

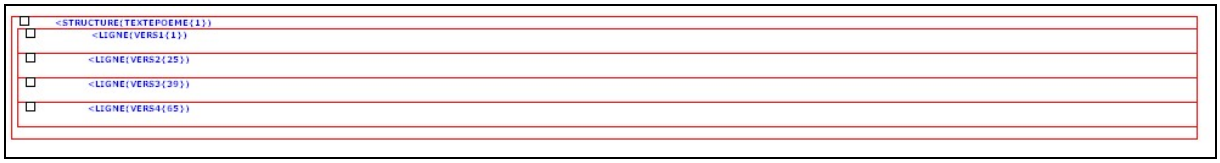
```
<?xml version="1.0" encoding="iso-8859-1"?>
<text id="x.c.hr.ce.1950.en">
 <front>
  <div type="adoption">
   <p>
    <date>1950-11-04</date>
```

```
    <name type="place">Rome</name>
   </p>
  </div>
  <div type="update">
   <p id="ce.1950.en_U.P">The text of the Convention had been
   amended according to the provisions of Protocol No. 3 (ETS
   No. 45), which entered into force on 21 September 1970, of
   Protocol No. 5 (ETS No. 55), which entered into force on 20
   December 1971 and of Protocol No. 8 (ETS No. 118), which
   entered into force on 1 January 1990, and comprised also the
   text of Protocol No. 2 (ETS No. 44) which, in accordance with
   Article 5, paragraph 3 thereof, had been an integral part of
   the Convention since its entry into force on 21 September
   1970. All provisions which had been amended or added by these
   Protocols are replaced by Protocol No. 11 (ETS No. 155), as
   from the date of its entry into force on 1 November 1998. As
   from that date, Protocol No. 9 (ETS No. 140), which entered
   into force on 1 October 1994, is repealed and Protocol No. 10
   (ETS No. 146) has lost its purpose.</p>
  </div>
 </front>
 <body>
  <head id="ce.1950.en_H">Convention for the Protection of Human
  Rights and Fundamental Freedoms as amended by Protocol No.
  11</head>
  <div type="preamble">
   <list type="simple">
    <item id="ce.1950.en_B.i-1">
     <seg type="membrestates">The governments signatory
     hereto, being members of the Council of Europe,</seg>
    </item>
   </list>
  </div>
 </body>
</text>
```

Le texte contient une ***Trame*** et un ***Cadre*** (chaque partie est parfaitement délimitée : balises de début et de fin explicites). Le ***Trameur*** peut travailler sur la ***Trame*** du texte et il est possible d'accéder au texte par son ***Cadre***.

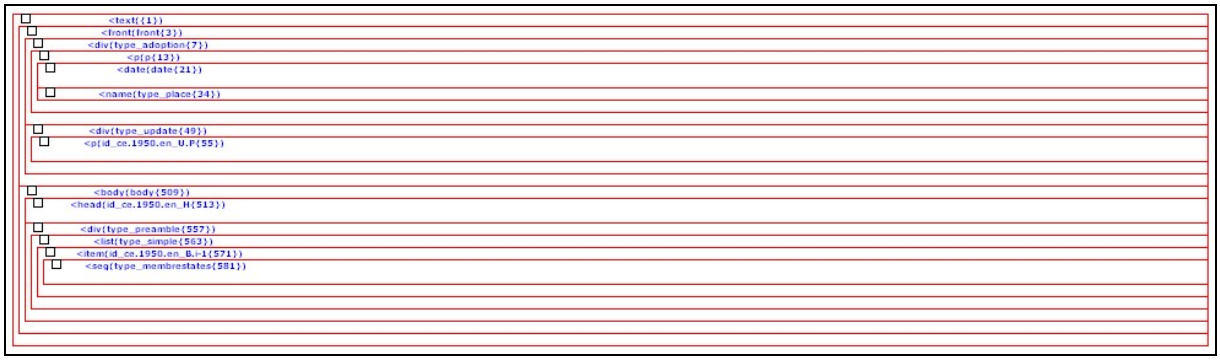
Pour les trois exemples précédents nous donnons à voir la représentation graphique construite par le ***Trameur*** du ***Cadre*** (complet) de chacun de ces textes :

Exemple n°1 :



Exemple n°2 :

Exemple n°3 :



9 Les dictionnaires

9.1 Les formes

Le chargement de la base déclenche l'insertion des dictionnaires des formes, des lemmes, des catégories (issues de l'étiquetage avec *Treetagger*) respectivement dans l'onglet **FORME-LEMME** et dans l'onglet **CATEGORIE-TAG.**

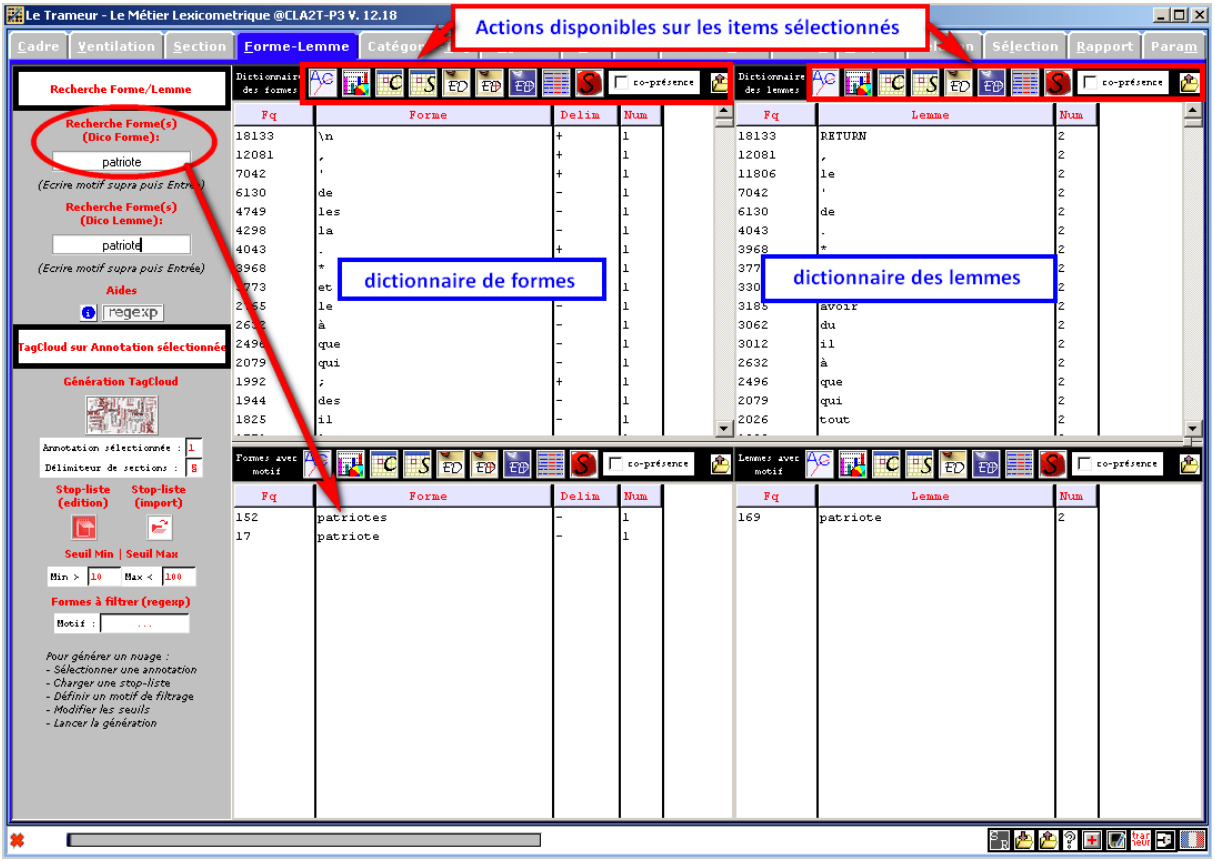


Figure 6 : *Le Trameur* Onglet FORME

Après sélection d'une forme (ou de plusieurs) dans le dictionnaire, il est possible de déclencher différentes actions :

Figure 7 : *Le Trameur* Actions sur Items sélectionnés à partir des dictionnaires

la représentation de la courbe d'accroissement du vocabulaire sélectionné (les formes choisies dans le dictionnaire)

la ventilation de cette forme (ou de plusieurs) dans la partition sélectionnée (*cf* onglet **VENTILATION**).

la ventilation de cette forme (ou plusieurs) sur une sélection de parties selectionnées via le *Gestionnaire de Sélection* (*cf* onglet **SELECTION**).

la recherche de cette forme (ou de plusieurs) dans la représentation cartographique du ***Cadre*** *infra* (*cf* onglet **CADRE**).

la recherche de cette forme (ou de plusieurs) dans la représentation cartographique de la carte des sections *infra* (*cf* onglet **SECTION**). Si l'option « co-présence » est activée, les sections mises au jour seront celles contenant l'ensemble des formes visées (ET logique)

l'édition de la forme visée pour éventuellement la corriger (si la forme est mise à jour, tous les items de la ***Trame*** ayant cette forme seront modifiés).

l'ajout d'un niveau d'un niveau d'annotation sur la forme visée (si la forme est mise à jour, tous les items de la ***Trame*** ayant cette forme seront modifiés pour cet ajout d'annotation)

la correction des différents niveaux d'annotation sur la forme visée (si la forme est mise à jour, tous les items de la ***Trame*** ayant cette forme seront modifiés pour ces modifications sur tous les niveaux d'annotation)

une concordance de cette forme (ou de plusieurs) (*cf* onglet **CONCORDANCE**).

la sélection de cette forme (ou de ces formes) *via* ses positions sur la ***Trame*** (*cf* onglet **SELECTION**).

Ces différentes opérations sont détaillées *infra*.

Les opérations similaires sont disponibles pour les lemmes, les catégories et les annotations importées.

On peut rechercher des formes dans le dictionnaire en définissant un motif de recherche dans la zone de saisie « *Recherche Forme(s)* ». Les formes trouvées seront listées au bas de l'onglet **FORME.** Les opérations décrites *supra* sur les formes sélectionnées sont là encore disponibles.

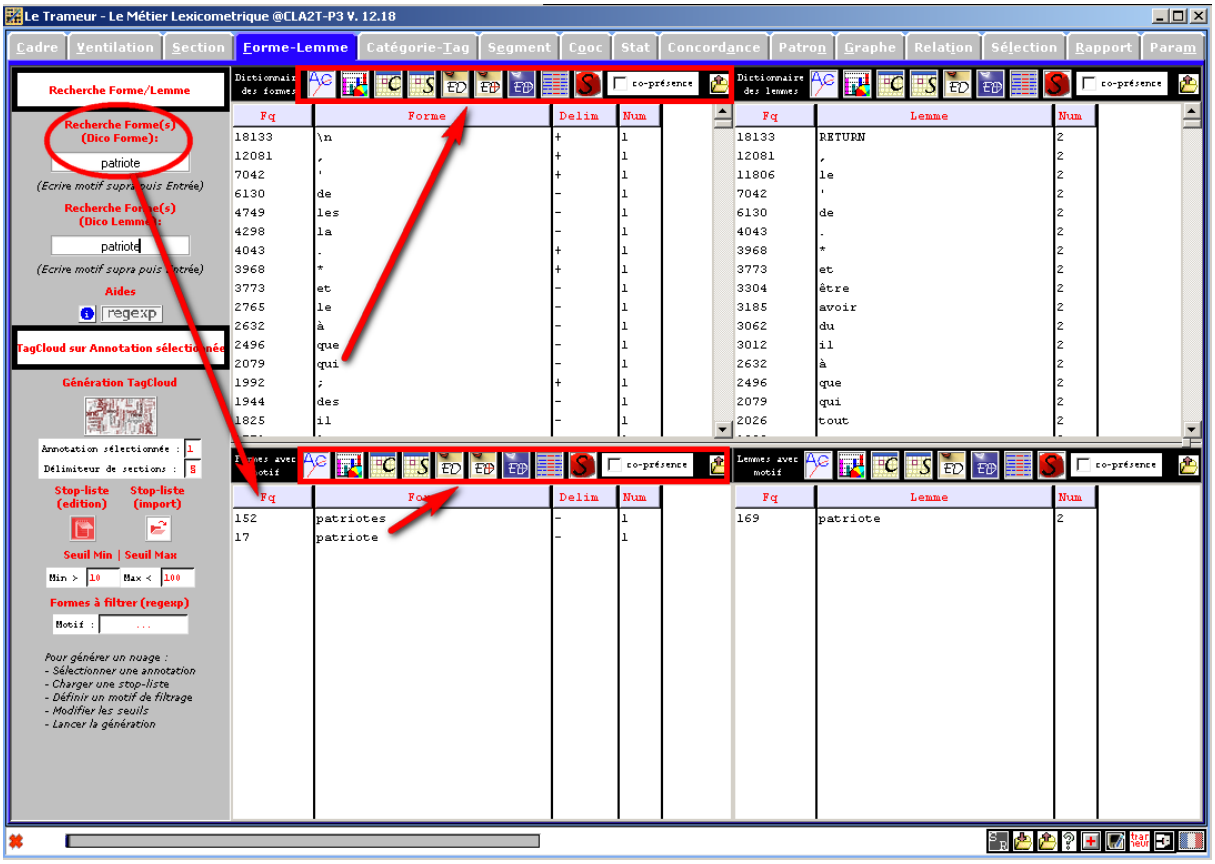


Figure 8 : *Le Trameur* Onglet FORME

9.2 Les lemmes

(*idem supra*)

9.3 Les catégories

(*idem supra*)

9.4 Les annotations importées

(*cf infra*)

10 Accès à la Trame par le Cadre

10.1 Mise au jour de la Trame

Dans les figures qui suivent, le fichier de travail utilisé est le *Dormeur du Val* dont on présente ci-dessous un extrait :

```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
```

La ***Trame*** du texte correspond à toutes les positions des **items** (délimiteurs et formes graphiques). Sur cet exemple :

En position 1, le caractère « espace » (considéré comme un délimiteur), en position 2, la forme graphique « Le », en position 3, le caractère « espace » etc.

Le ***Trameur*** produira une segmentation que l'on peut schématiser ainsi sur le début du texte :

Forme		Le		dormeur		du		val	RC		C	…
Positions	1	2	3	4	5	6	7	8	9	10	11	…

(RC = retour à la ligne)

Le ***Cadre*** du texte est matérialisé ici par des balises indiquant le début d'une partie (sous la forme `<partie=valeur>`). Ce balisage traduit une structuration en parties du texte (1 titre et 1 poème contenant des vers). Plus précisément, la partie `STRUCTURE=TITRE` recouvre les items de la position 1 à la position 8 et la partie `STRUCTURE=POEME` couvre les items à partir de la position 9 jusqu'à la dernière. On verra *infra* qu'il est possible de décrire le ***Cadre*** d'un texte de manière plus formalisée *via* XML.

Après chargement de la base utilisée, le ***Trameur*** a l'allure suivante :

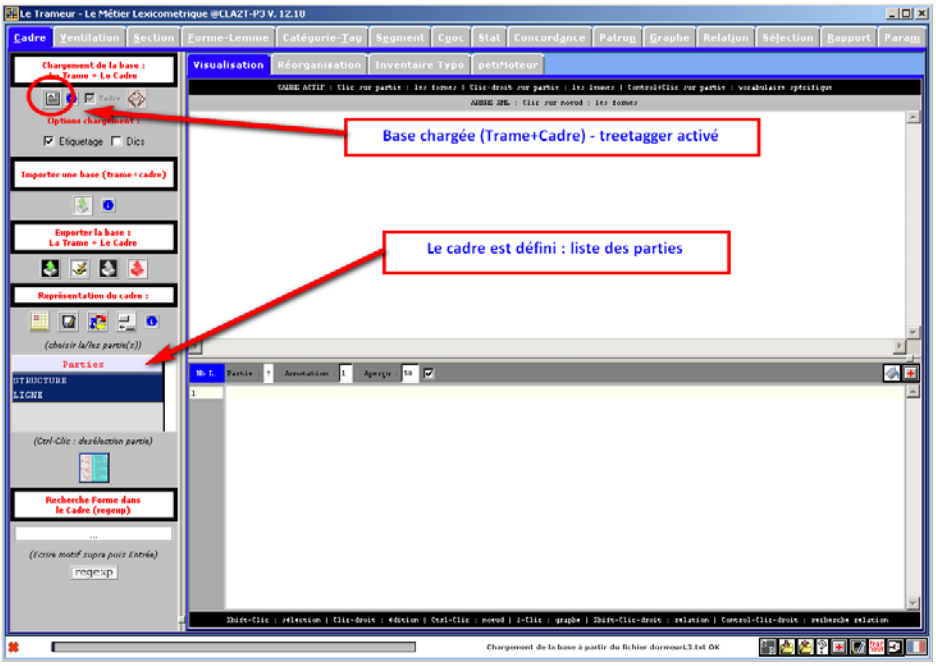


Figure 9 : *Le Trameur* Onglet CADRE, base chargée

Les différentes parties constituant le ***Cadre*** sont visibles dans la zone ***Parties***. Pour visualiser tout ou partie du ***Cadre***, il convient de sélectionner une ou plusieurs parties, puis d'activer le bouton . Après génération du ***Cadre*** (dans notre exemple, 2 parties ont été sélectionnées), le ***Trameur*** a l'allure suivante :

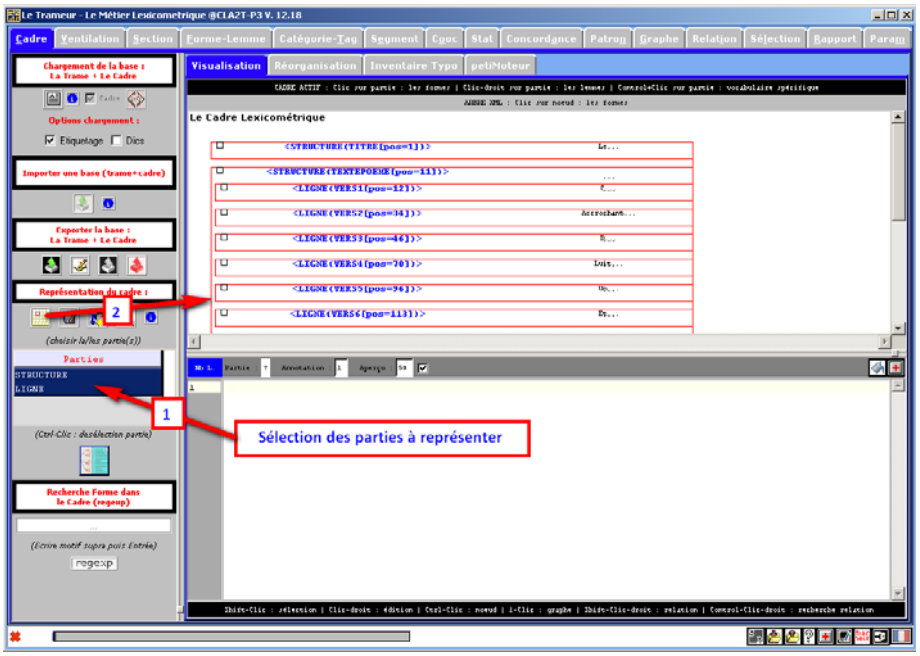


Figure 10 : *Le Trameur* Onglet CADRE, *Cadre* généré

Une représentation graphique des parties constituant le texte est construite dans la *zone d'édition du **Cadre***. Cette représentation donne à voir l'emboîtement éventuel des différentes parties : dans notre exemple, les VERS sont contenus dans la partie POEME. En cliquant (*via* un *clic-gauche*) sur le nom d'une partie, on active l'affichage de son contenu textuel (*les formes i.e les mots*) dans la *zone d'édition de la **Trame*** au bas de l'onglet **CADRE**. Dans la figure suivante, on donne à voir le contenu de la partie `STRUCTURE=TEXTEPOEME` :

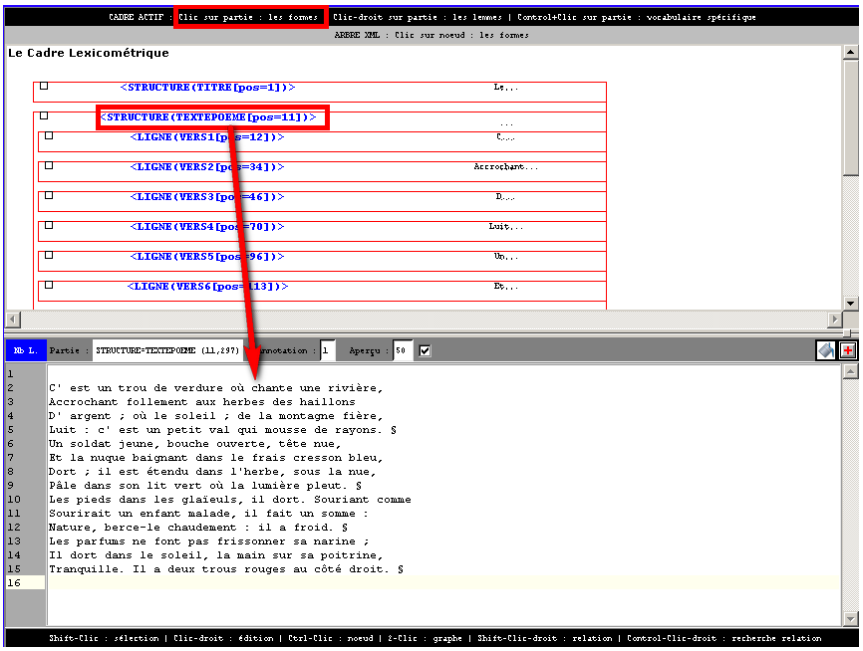


Figure 11 : *Le Trameur* Onglet CADRE, contenu textuel d'une partie (les mots)

En cliquant (*via* un *clic-droit*) sur le nom d'une partie, on active l'affichage de son contenu textuel (*les lemmes*) dans la zone d'édition au bas de l'onglet **CADRE**. Dans la figure suivante, on donne à voir le contenu de la partie `STRUCTURE=POEME` :

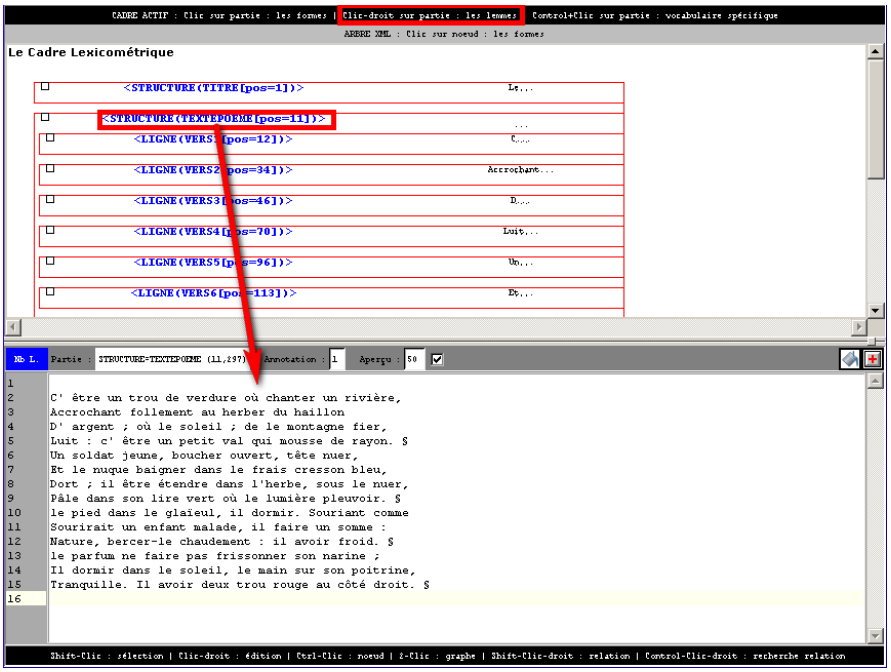


Figure 12 : *Le Trameur* Onglet CADRE, contenu textuel d'une partie (les lemmes)

Si la base à charger est au format XML, on peut aussi construire une représentation arborescente du fichier chargé.

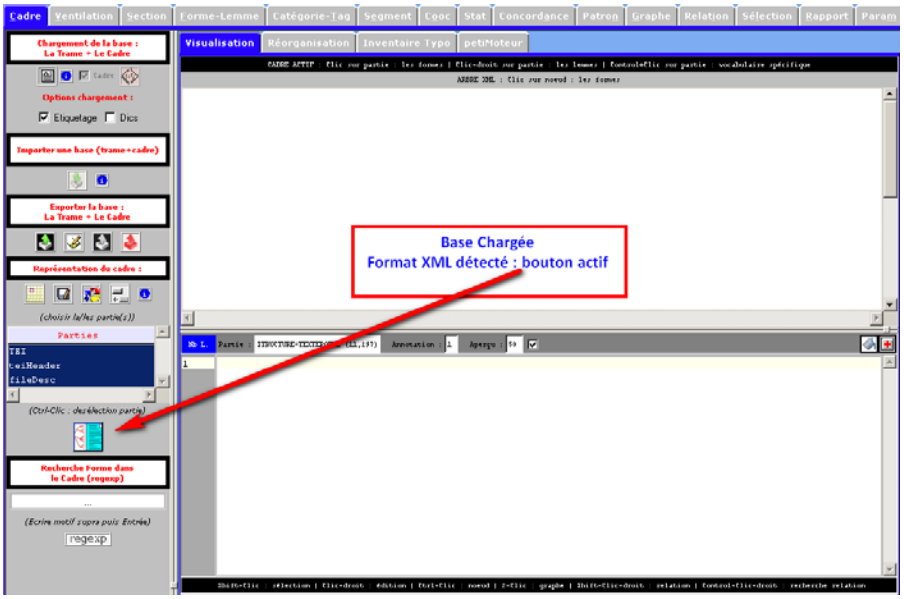


Figure 13 : *Le Trameur* Onglet CADRE, chargement d'une représentation arborescente d'une base au format XML (1)

Par défaut, le bouton est inactif ; dans le cas d'un fichier au format XML, il devient actif et permet le déclenchement de la représentation arborescente du fichier :

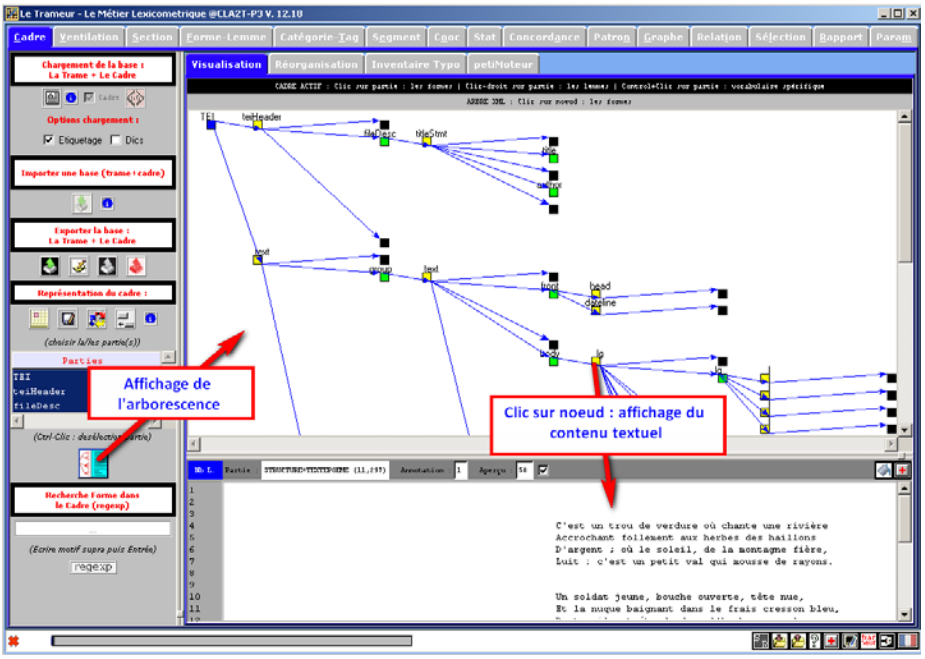


Figure 14 : *Le Trameur* Onglet CADRE, chargement d'une représentation arborescente d'une base au format XML (2)

10.2 Choix du niveau de l'annotation à afficher

Une zone de saisie disponible dans l'éditeur du **CADRE** et dans l'éditeur de **SECTION** permet de sélectionner le numéro d'annotation à utiliser pour afficher la partie ou la section visée. Rappel : par convention, les annotations associées aux items de la ***Trame*** sont typées de la manière suivante :

- le numéro d'annotation n°1 correspond à « Forme »
- le numéro d'annotation n°2 correspond à « Lemme »
- le numéro d'annotation n°3 correspond à « Catégorie »
- Toute autre annotation importée aura un numéro supérieur ou égal à 4.

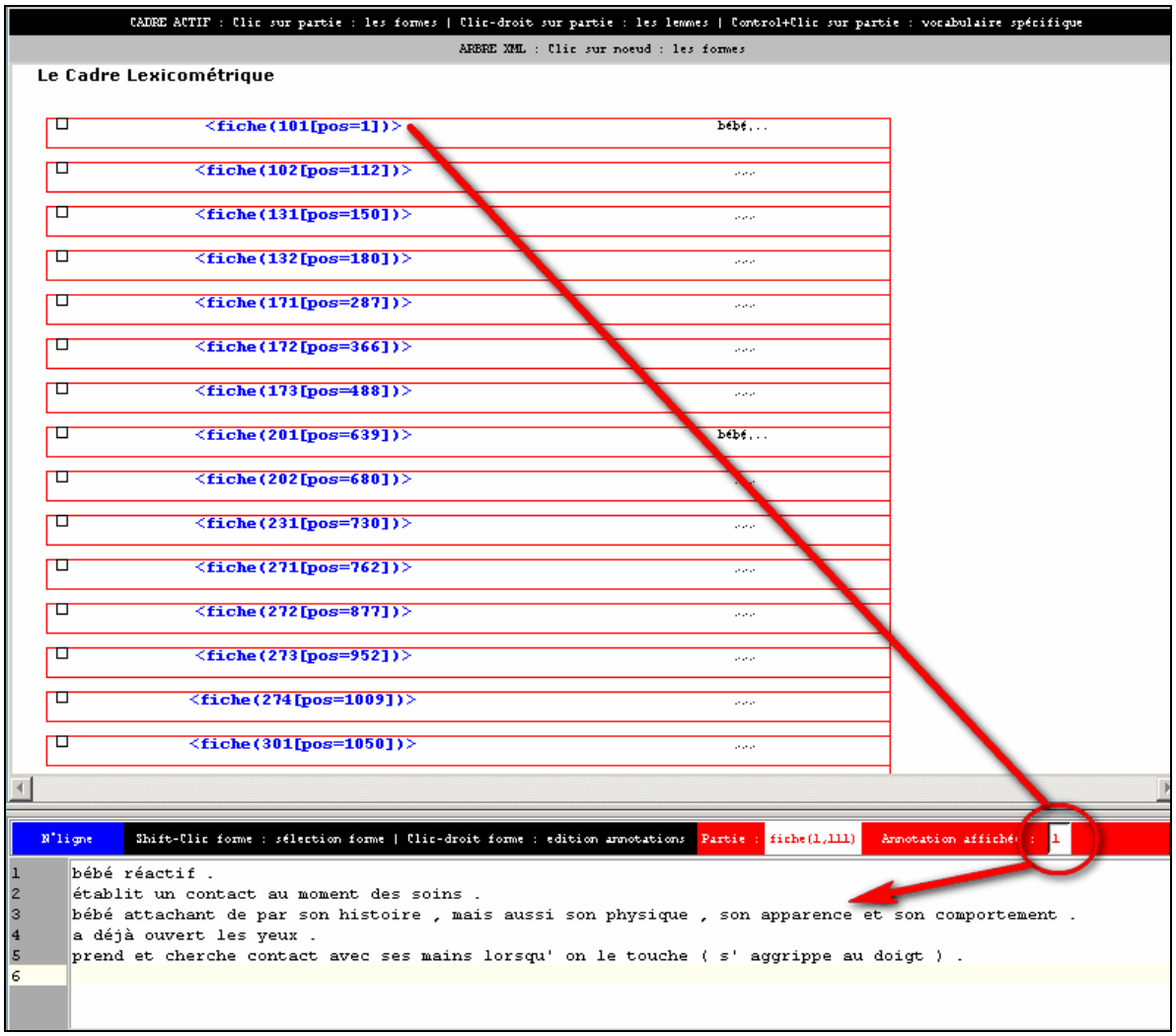


Figure 15 : *Le Trameur* Affichage d'une partie du CADRE (annotation affichée = 1)

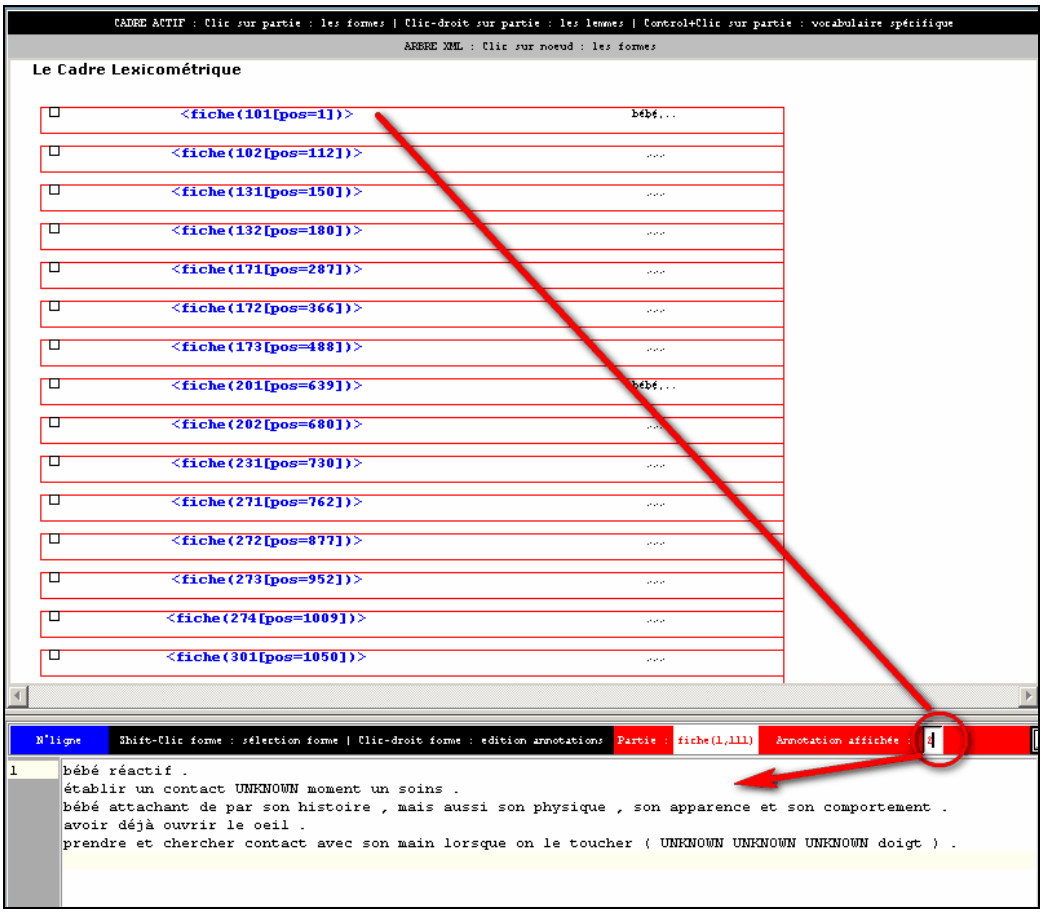


Figure 16 : *Le Trameur* Affichage d'une partie du CADRE (annotation affichée = 2)

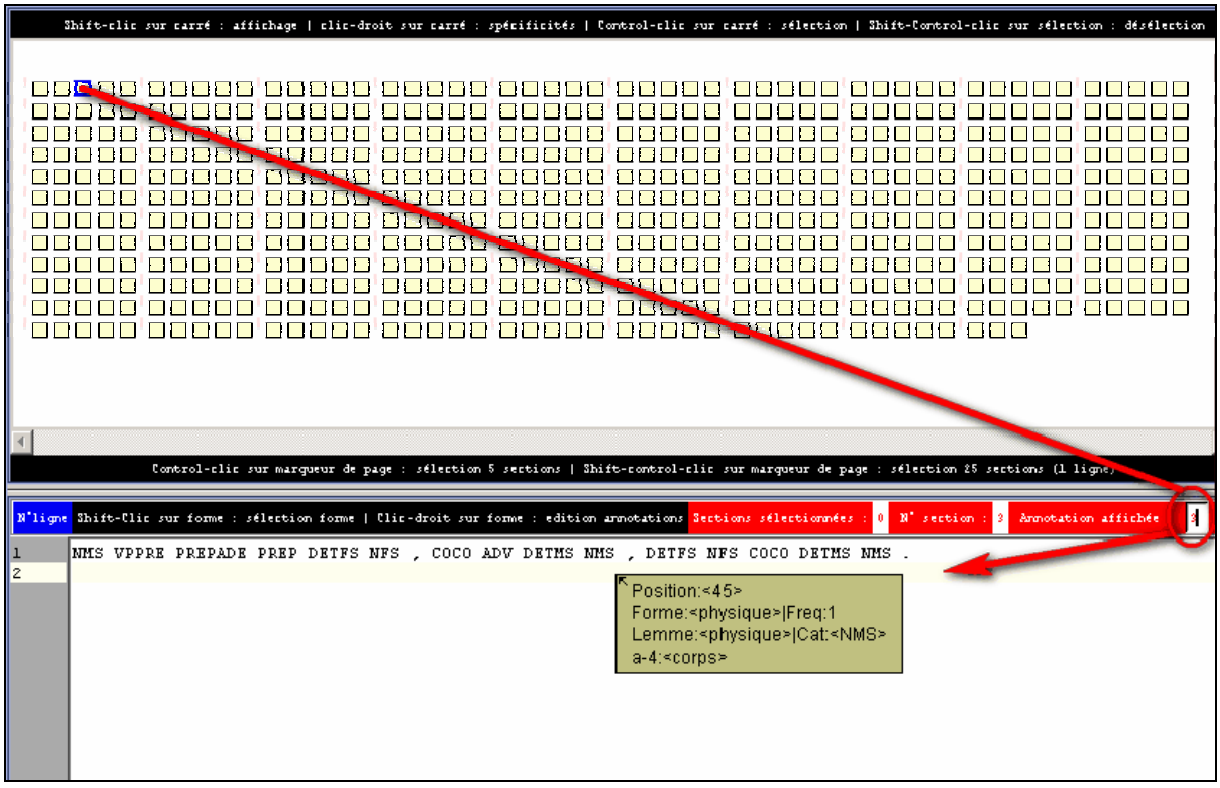


Figure 17 : *Le Trameur* Affichage d'une section (annotation affichée = 3)

10.3 Recherche de chaînes

La figure suivante illustre une des fonctionnalités classiques d'un éditeur que l'on retrouve dans le ***Trameur*** : la recherche de chaîne de caractères.

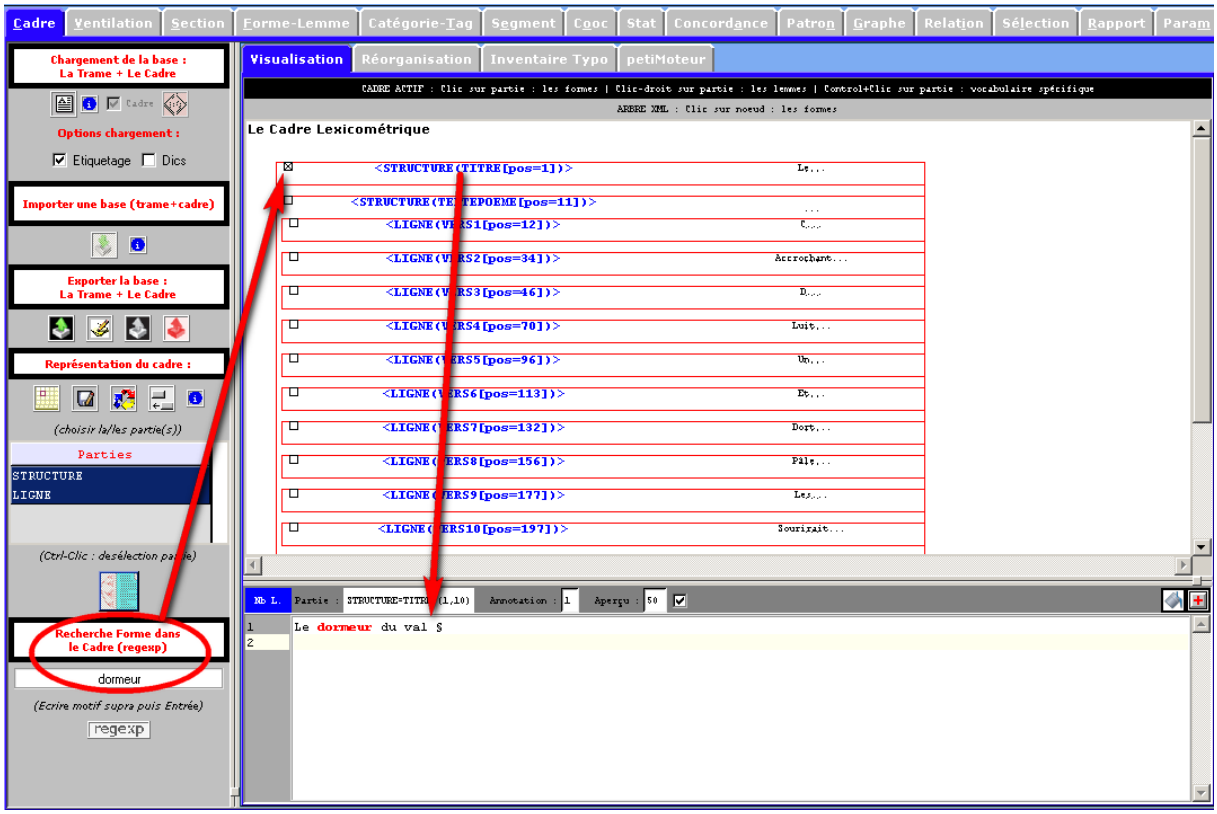


Figure 18 : *Le Trameur* Onglet CADRE, recherche de chaînes

On dispose dans l'onglet **CADRE** d'une zone de saisie permettant de lancer des recherches de chaînes : "*Recherche*". Pour rechercher une forme, il convient de l'insérer dans la zone de saisie puis de lancer la recherche en tapant la touche « Entrée ».

Dans la figure précédente, la forme cherchée (dormeur) est présente dans la partie `STRUCTURE(TITRE)` : son contenu textuel est visible *via* un *clic-gauche* sur le carré associé à la partie (la croix intégrée au carré indiquant la présence de la forme visée).

On peut aussi lancer une recherche en sélectionnant une forme dans les dictionnaires disponibles (et *via* le raccourci clavier défini).

10.4 Recherche « incrémentale » de chaînes

On peut aussi activer un mode de recherche dit *incrémentale*. Le raccourci clavier *Control-s* active ce mode de recherche permettant de définir *incrémentalement* un motif de recherche.

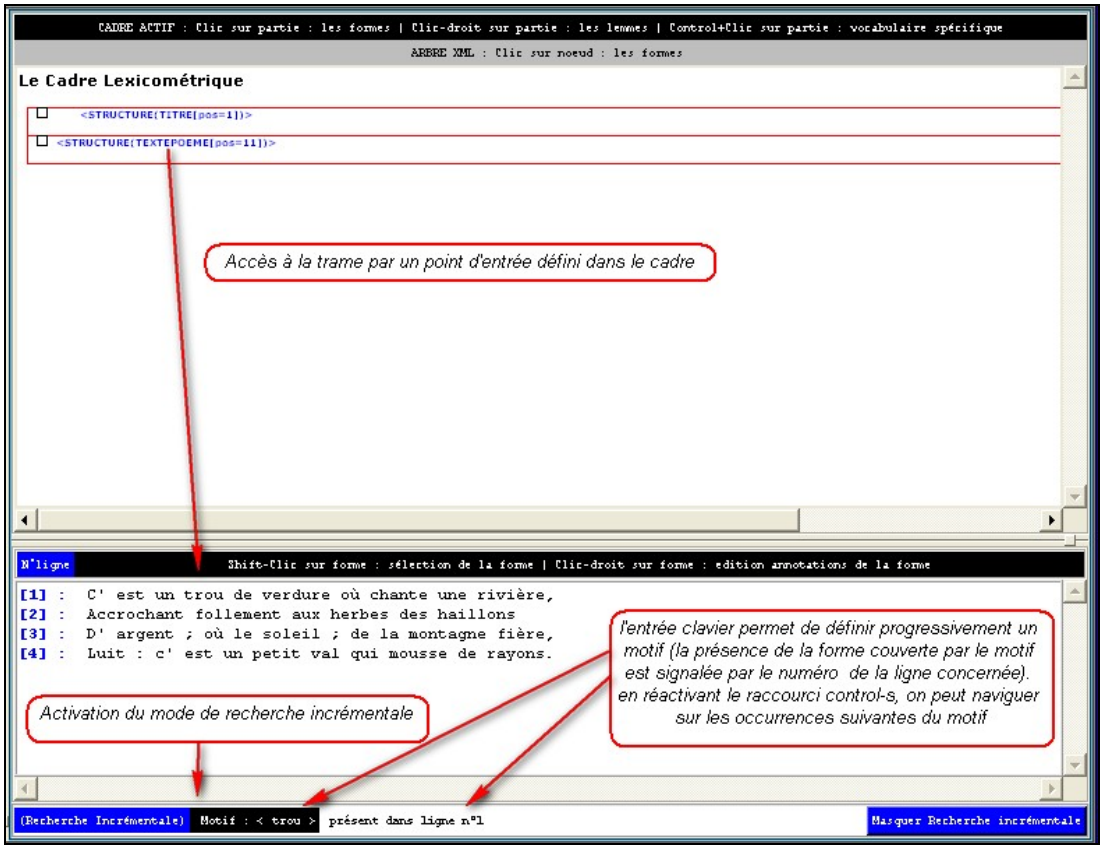

CADRE ACTIF : Clic sur partie : les formes | Clic-droit sur partie : les lemmes | Control+Clic sur partie : vocabulaire spécifique
ARBRE XML : Clic sur noeud : les formes
Le Cadre Lexicométrique
<STRUCTURE(TITRE[pos=1])>
<STRUCTURE(TEXTEPOEME[pos=11])>
Accès à la trame par un point d'entrée défini dans le cadre
N°ligne
Shift-Clic sur forme : sélection de la forme | Clic-droit sur forme : edition annotations de la forme
[1] : C' est un trou de verdure où chante une rivière,
[2] : Accrochant follement aux herbes des haillons
[3] : D' argent ; où le soleil ; de la montagne fière,
[4] : Luit : c' est un petit val qui mousse de rayons.
Activation du mode de recherche incrémentale
l'entrée clavier permet de définir progressivement un motif (la présence de la forme couverte par le motif est signalée par le numéro de la ligne concernée). en réactivant le raccourci control-s, on peut naviguer sur les occurrences suivantes du motif
(Recherche Incrémentale)
Motif : < trou >
présent dans ligne n°1
Masquer Recherche incrémentale

Figure 19 : *Le Trameur* Onglet CADRE, recherche incrémentale de chaînes

10.5 Visualisation des différentes annotations des items de la *Trame*

Il est possible de visualiser les différentes annotations associées aux items présents dans la zone d'édition (toute ou une partie de la ***Trame***). Il suffit de placer la souris sur une des formes graphiques présentes dans cette zone d'édition (mot ou lemme), une boîte d'édition contextuelle apparaît et donne des précisions sur la forme visée : forme graphique, fréquence, lemme, catégorie (et annotation complémentaire si disponible).

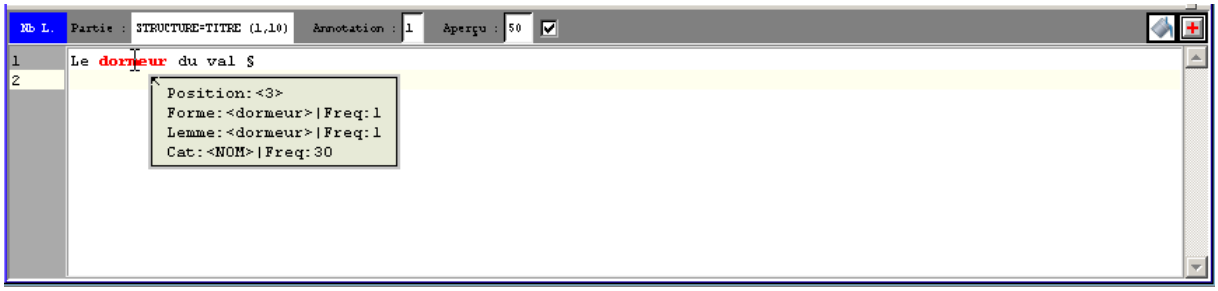


Figure 20 : *Le Trameur* Onglet CADRE, visualisation détaillée d'un item

10.6 Edition et modification des items de la Trame

Dans la même zone d'édition (forme ou lemme), il est possible d'éditer et de modifier les contenus de chaque item. Un *clic-droit* sur un item (forme ou lemme) déclenche l'apparition d'une fenêtre d'édition de l'item visé dans laquelle il est possible de modifier sa forme, son lemme ou sa catégorie (et toute annotation importée) :

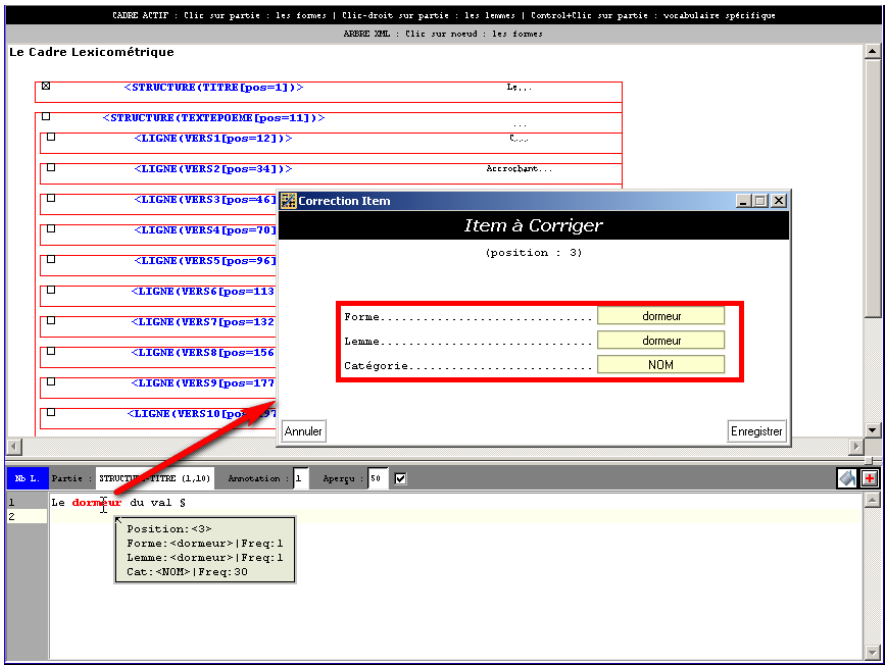


Figure 21 : *Le Trameur* Onglet CADRE, édition d'un item

Après enregistrement des modifications réalisées, la fenêtre d'édition initiale est rechargée en tenant compte des modifications apportées :

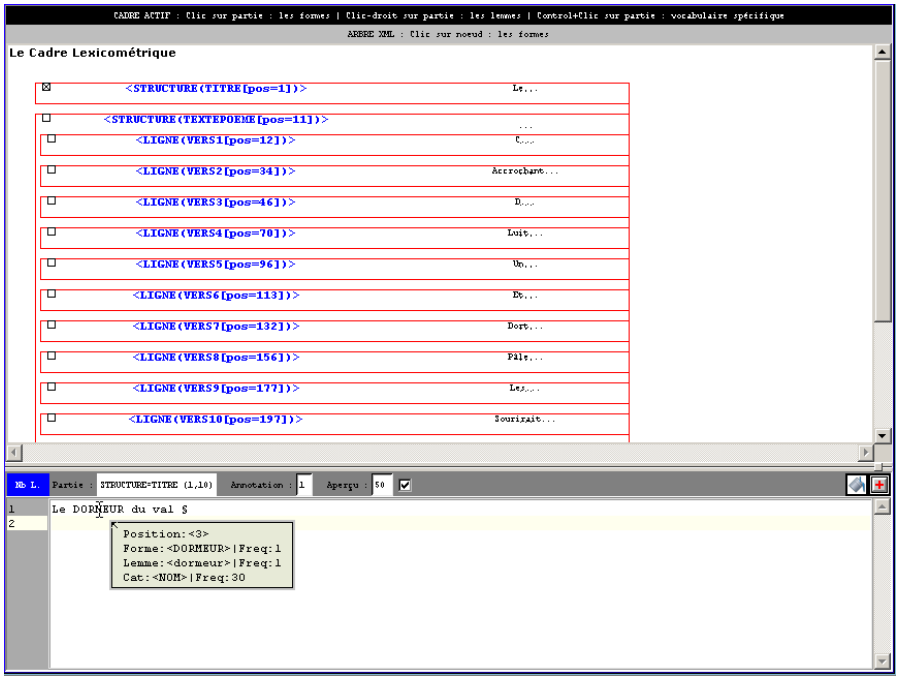
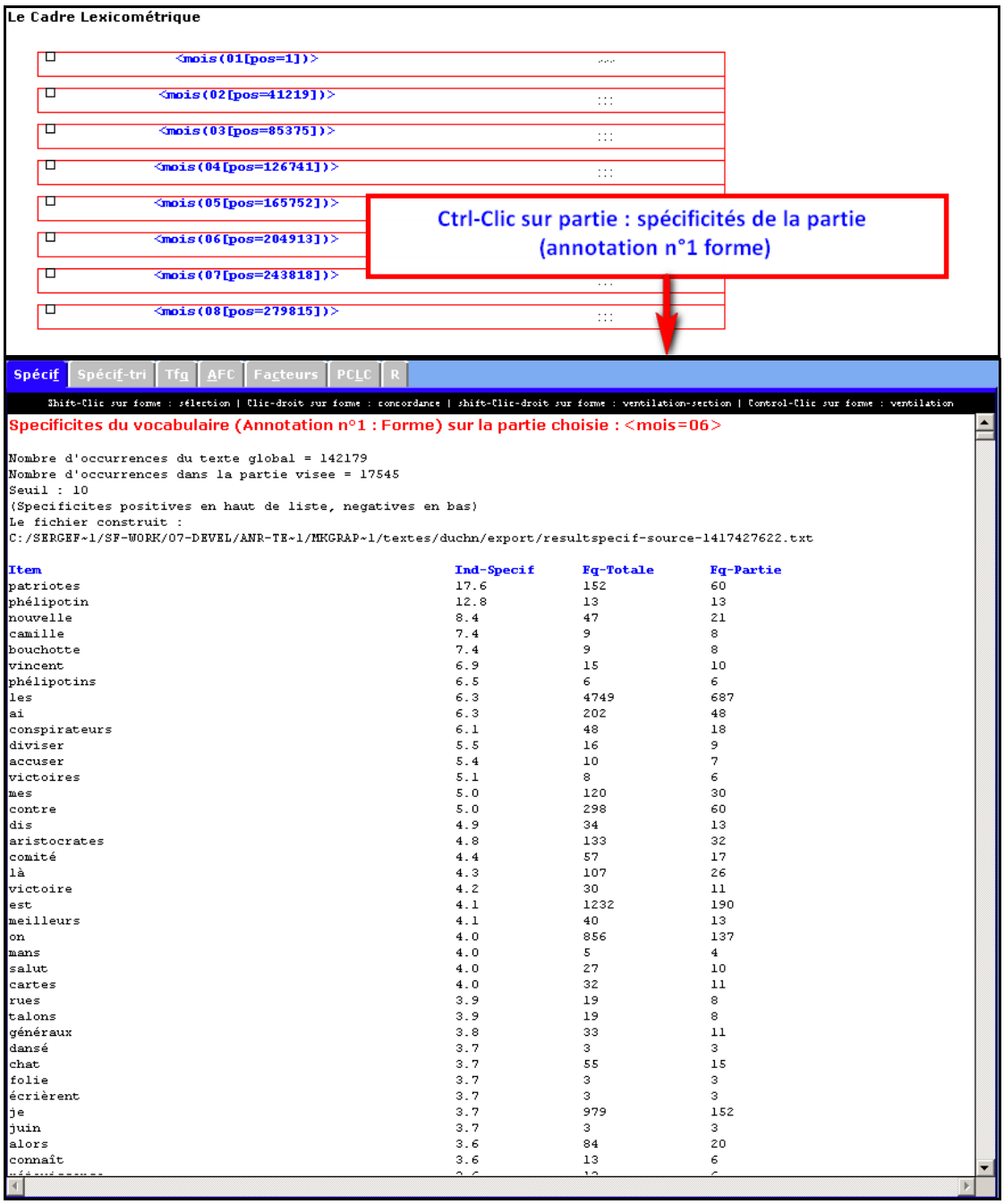


Figure 22 : *Le Trameur* Onglet CADRE, modification d'un item

Et les différents dictionnaires impactés par ces modifications sont rechargés.

10.7 Le vocabulaire spécifique d'une partie

Le ***Cadre*** dispose d'une fonctionnalité supplémentaire pour agir sur les contenus textuels des parties repérées. On a vu *supra* qu'un *clic-gauche* sur une partie déclenchait l'affichage du contenu textuel associé à cette partie dans la zone d'édition prévue à cet effet. Le raccourci clavier *control+clic-gauche* sur une partie déclenche le calcul du vocabulaire spécifique[7] de la partie visée.



Item	Ind-Specif	Fq-Totale	Fq-Partie
patriotes	17.6	152	60
phélipotin	12.8	13	13
nouvelle	8.4	47	21
camille	7.4	9	8
bouchotte	7.4	9	8
vincent	6.9	15	10
phélipotins	6.5	6	6
les	6.3	4749	687
ai	6.3	202	48
conspirateurs	6.1	48	18
diviser	5.5	16	9
accuser	5.4	10	7
victoires	5.1	8	6
mes	5.0	120	30
contre	5.0	298	60
dis	4.9	34	13
aristocrates	4.8	133	32
comité	4.4	57	17
là	4.3	107	26
victoire	4.2	30	11
est	4.1	1232	190
meilleurs	4.1	40	13
on	4.0	856	137
mans	4.0	5	4
salut	4.0	27	10
cartes	4.0	32	11
rues	3.9	19	8
talons	3.9	19	8
généraux	3.8	33	11
dansé	3.7	3	3
chat	3.7	55	15
folie	3.7	3	3
écrièrent	3.7	3	3
je	3.7	979	152
juin	3.7	3	3
alors	3.6	84	20
connaît	3.6	13	6

Figure 23 : *Le Trameur* Onglet CADRE, le vocabulaire spécifique d'une partie

[7] Cf Partie « Vocabulaire spécificique » pour le paramétrage de ce calcul

11 La carte des sections : représentation cartographique de la *Trame*

L'onglet **SECTION** permet de donner à voir une représentation graphique du texte sous la forme d'une carte de sections (identique à celle produite par **Lexico3**[8] *via* l'outil "*Carte des Sections*[9]"). Le bouton permet de construire cette représentation cartographique sur la base du caractère délimiteur choisi pour délimiter les sections.

11.1 La carte des sections : une vue globale de la Trame

La ***carte des sections*** permet une visualisation du corpus découpé en sections par la promotion d'un (ou de plusieurs) caractère particulier (paragraphes, point, etc.) au statut de ***délimiteur de section***. La carte des sections recouvre ainsi intégralement la ***Trame*** textuelle (**une vue globale de la *Trame***).

Si on considère les 2 exemples ci-dessous, dans chacun d'eux le caractère § est intégré dans le texte pour marquer des zones.

Dans le premier exemple, le caractère § induit la présence de 6 sections.

Dans le second, on retrouve aussi six sections, mais la ***Trame*** textuelle dispose en outre d'une ***Cadre*** (des jalons du type « balise » organisant le texte en parties)

```
1                    Le dormeur du val §
2
3   C'est un trou de verdure où chante une rivière,
4  Accrochant follement aux herbes des haillons
5  D'argent ; où le soleil, de la montagne fière,
6  Luit : c'est un petit val qui mousse de rayons. §
7
8   Un soldat jeune, bouche ouverte, tête nue,
9  Et la nuque baignant dans le frais cresson bleu,
10 Dort ; il est étendu dans l'herbe, sous la nue,
11 Pâle dans son lit vert où la lumière pleut. §
12
13  Les pieds dans les glaïeuls, il dort. Souriant comme
14 Sourirait un enfant malade, il fait un somme :
15 Nature, berce-le chaudement : il a froid. §
16
17  Les parfums ne font pas frissonner sa narine ;
18 Il dort dans le soleil, la main sur sa poitrine,
19 Tranquille. Il a deux trous rouges au côté droit. §
20
21                Arthur Rimbaud §
```

Figure 24 : *Le Trameur* Onglet SECTION, texte brut avec caractère « délimiteur de sections »

[8] http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/
[9] http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/manuelsL3/manuel-3.41.htm#_Toc34045613

```
<STRUCTURE="TITRE">Le dormeur du val §
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1">C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2">Accrochant follement aux herbes des haillons
<LIGNE="VERS3">D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4">Luit : c' est un petit val qui mousse de rayons. §
<LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
<LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
<LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
<LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut. §
<LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
<LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
<LIGNE="VERS11">Nature, berce-le chaudement : il a froid. §
<LIGNE="VERS12">Les parfums ne font pas frissonner sa narine ;
<LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
<LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit. §
<STRUCTURE="AUTEUR">Arthur Rimbaud §
```

Figure 25 : *Le Trameur* Onglet SECTION, texte structuré avec caractère « délimiteur de sections »

Pour ces deux exemples le caractère délimiteur § est utilisé dans les figures suivantes pour construire la carte des sections associée :

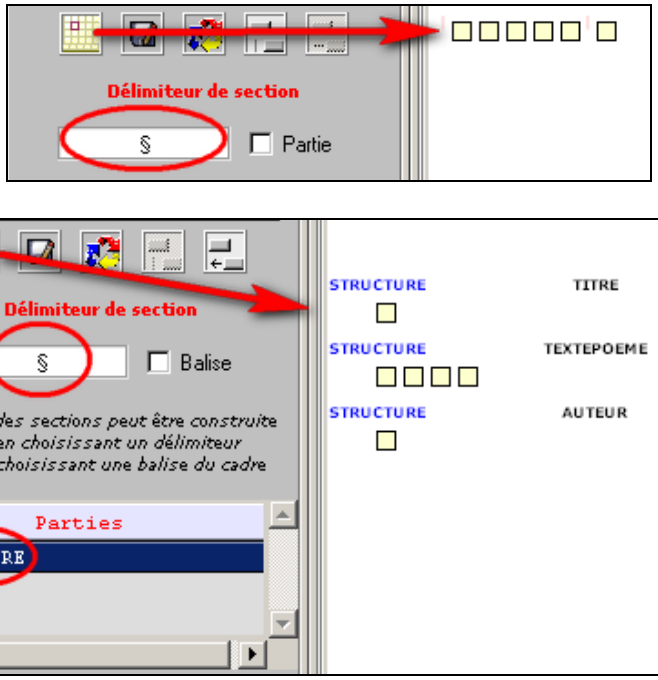


Figure 26 : *Le Trameur* Onglet SECTION, construction de la carte des sections (3)

La seconde figure illustre aussi la possibilité d'introduire des éléments du ***Cadre*** dans la carte des sections.

En cliquant sur un des carrés, on active l'affichage de son contenu textuel dans la zone d'édition au bas de l'onglet **SECTION**.

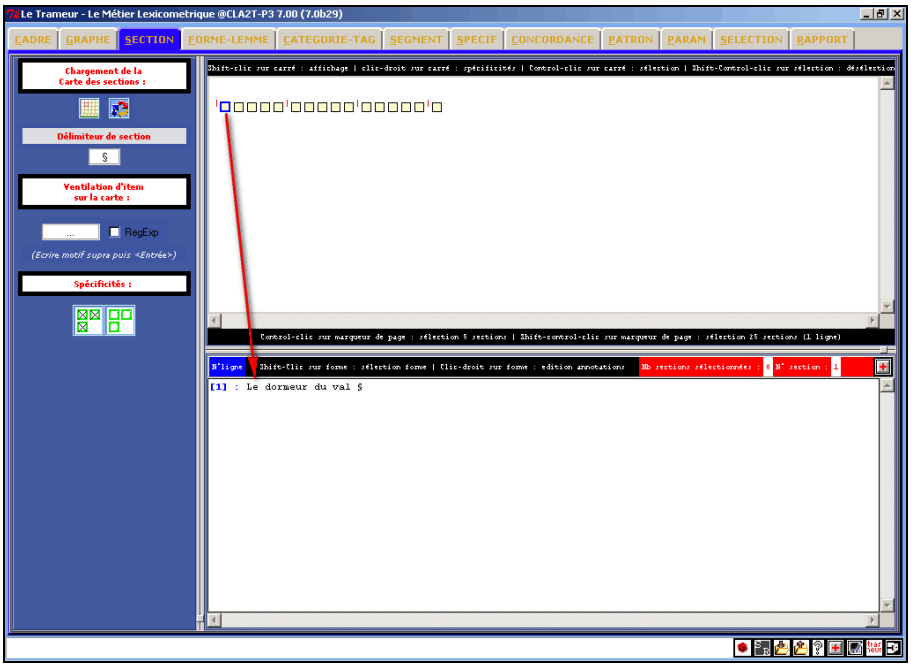


Figure 27 : *Le Trameur* Onglet SECTION, édition de la *Trame* par une section

Paramétrage de l'affichage de la carte des sections

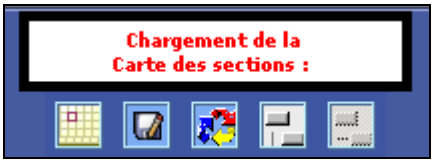


Figure 28 : *Le Trameur* Onglet SECTION, paramétrage de l'affichage

Le bouton active le chargement de la carte. Par défaut, cet affichage insère linéairement les sections et sans alignement à gauche après insertion d'un marqueur de partie. De plus, un séparateur de blocs de sections est inséré après affichage de 5 sections : ce séparateur est utilisable ensuite pour sélectionner les sections (*cf infra*).

- Alignement à gauche

Ce bouton est inhibé par défaut, son activation (*clic-droit*) permet de représenter la carte en alignant les sections à gauche après chaque marquage de partie. Désactivation de cette fonction par un *clic* sur le bouton

- Insertion d'un marqueur de bloc de sections

Ce bouton est actif par défaut, son inhibition permet de ne pas afficher le marqueur de blocs de sections.
Pour inhiber cette fonction : *clic-droit*
Pour l'activer : *clic*

On verra infra (*cf* Chantiers en cours) une illustration de l'utilisation de ces 2 derniers boutons.

11.2 La carte des sections : une vue partielle de la Trame

Si on considère cette représentation du même texte (ici, format TEI) :

```
<?xml version="1.0" encoding="iso-8859-1"?>
<?xml-stylesheet type="text/xsl" href="q2.xsl"?>
<TEI>
    <teiHeader>
        <fileDesc>
            <titleStmt>
                <title>Poésies</title>
                <author>Arthur Rimbaud</author>
            </titleStmt>
        </fileDesc>
    </teiHeader>
    <text>
        <group>
            <text>
                <front>
                    <head>Le Dormeur du Val</head>
                    <dateline>7 octobre 1870</dateline>
                </front>
                <body>
                    <lg>
                        <lg>
                            <l>C'est un trou de verdure où chante une rivière</l>
                            <l>Accrochant follement aux herbes des haillons</l>
                            <l>D'argent ; où le soleil, de la montagne fière,</l>
                            <l>Luit : c'est un petit val qui mousse de rayons.</l>
                        </lg>
                        <lg>
                            <l>Un soldat jeune, bouche ouverte, tête nue,</l>
                            <l>Et la nuque baignant dans le frais cresson bleu,</l>
                            <l>Dort ; il est étendu dans l'herbe, sous la nue,</l>
                            <l>Pâle dans son lit vert où la lumière pleut.</l>
                        </lg>
                        <lg>
                            <l>Les pieds dans les glaïeuls, il dort. Souriant comme</l>
                            <l>Sourirait un enfant malade, il fait un somme :</l>
                            <l>Nature, berce-le chaudement : il a froid.</l>
                        </lg>
                        <lg>
                            <l>Les parfums ne font pas frissonner sa narine ;</l>
                            <l>Il dort dans le soleil, la main sur sa poitrine</l>
                            <l>Tranquille. Il a deux trous rouges au côté droit.</l>
                        </lg>
                    </lg>
                </body>
            </text>
        </group>
    </text>
</TEI>
```

Figure 29 : *Le Trameur* Onglet SECTION, texte TEI sans caractère « délimiteur de sections » explicite

Dans cet exemple, aucun caractère n'est disponible explicitement pour mettre au jour de manière systématique des sections.

Cependant, certains éléments du ***Cadre*** permettent de délimiter des sections particulières :

- L'élément `<l>…</l>` permet de mettre au jour les vers (un à un) ; dans ce cas cet élément peut définir une section correspondant à un vers
- L'élément `<lg>…</lg>` permet de mettre au jour les strophes (une à une) : dans ce cas cet élément peut définir une section correspondant à une strophe.

La mise en œuvre dans le ***Trameur*** de cette prise en charge des éléments du ***Cadre*** pour construire la carte des sections se déroule de la manière suivante :

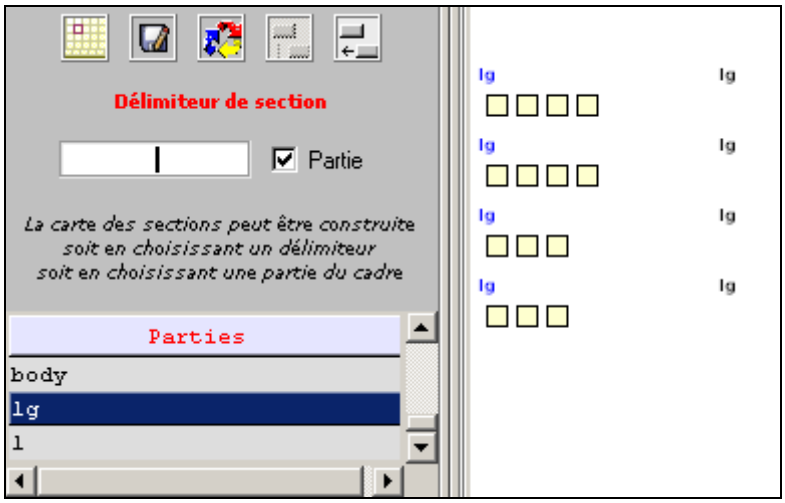


Figure 30 : *Le Trameur* Onglet SECTION, sections définies implicitement par un élement du *Cadre* (avec affichage d'une autre partie : `lg`)

Dans la figure précédente, la case à cocher « Partie » permet de spécifier que la construction de la carte utilise un élément du ***Cadre*** (une balise) ; le nom de la balise « `l` » à utiliser est inséré dans la zone de saisie. Comme précédemment, la partie lg est utilisée pour marquer certaines parties dans la carte construite.

Dans l'exemple qui suit, la balise utilisée pour la représentation de la carte est la balise `lg` :

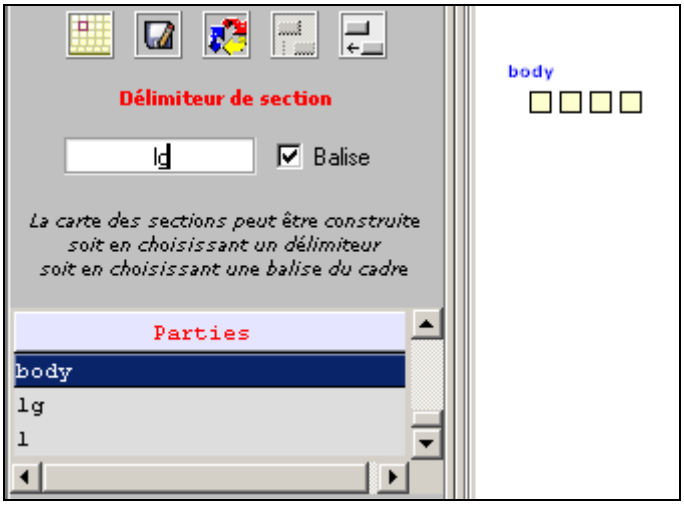


Figure 31 : *Le Trameur* Onglet SECTION, sections définies implicitement par un élement du *Cadre* (avec affichage d'une autre partie : `body`)

A la différence de la carte des sections définie précédemment, la carte des sections construite via un élément du ***Cadre*** offre ainsi **une vue partielle** de la ***Trame*** : celle induite par les différentes localisations des parties associées à l'élément sélectionné dans le ***Cadre***.

11.3 La carte des sections : masquage de sections

Quelque soit le mode sélectionné pour afficher la carte des sections (caractère délimiteur ou élement du cadre), il est possible de masquer temporairement l'affichage de sections pour une ou plusieurs parties données. Dans l'exemple qui suit, la carte des sections est initialement dessinée en affichant toutes les sections :

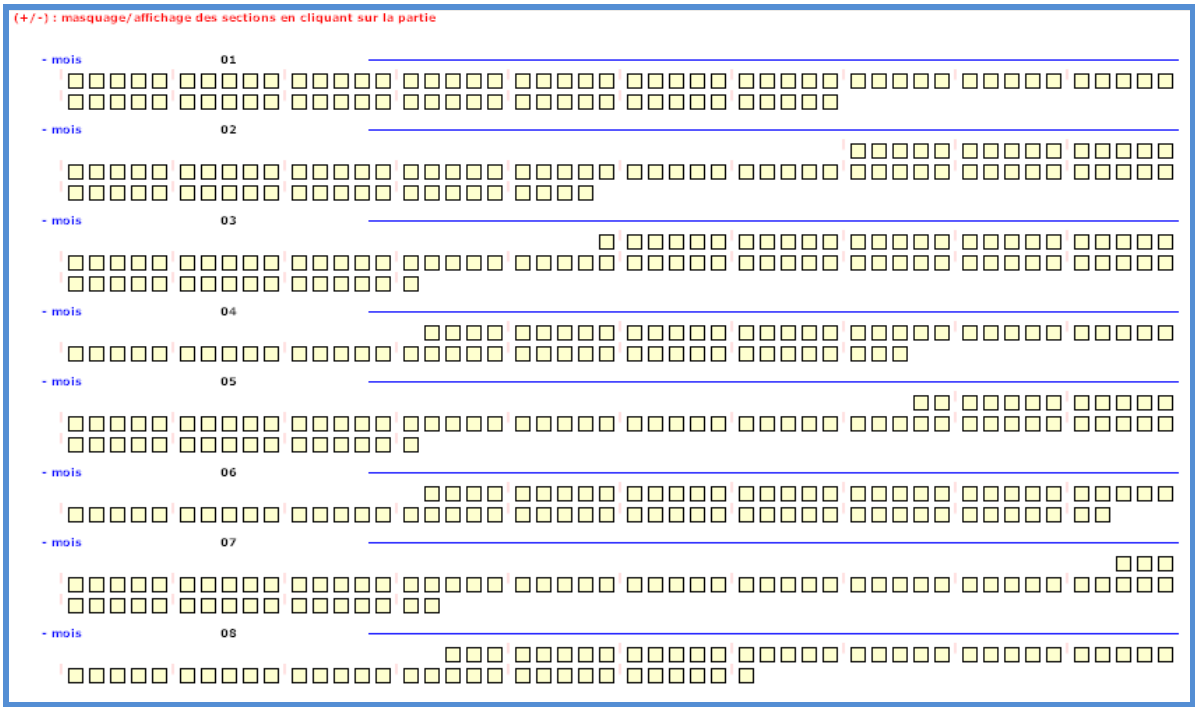


Figure 32 : *Le Trameur* Onglet SECTION, masquage de sections (1)

L'entête de la figure précédente mentionne la possibilité de masquer les sections des parties dans le nom est précédé du caractère + (réciproquement, d'afficher les sections des parties dont le nom est précédé du caractère -). La figure qui suit illustre cette fonctionnalité : toutes les parties sont masquées sauf deux.

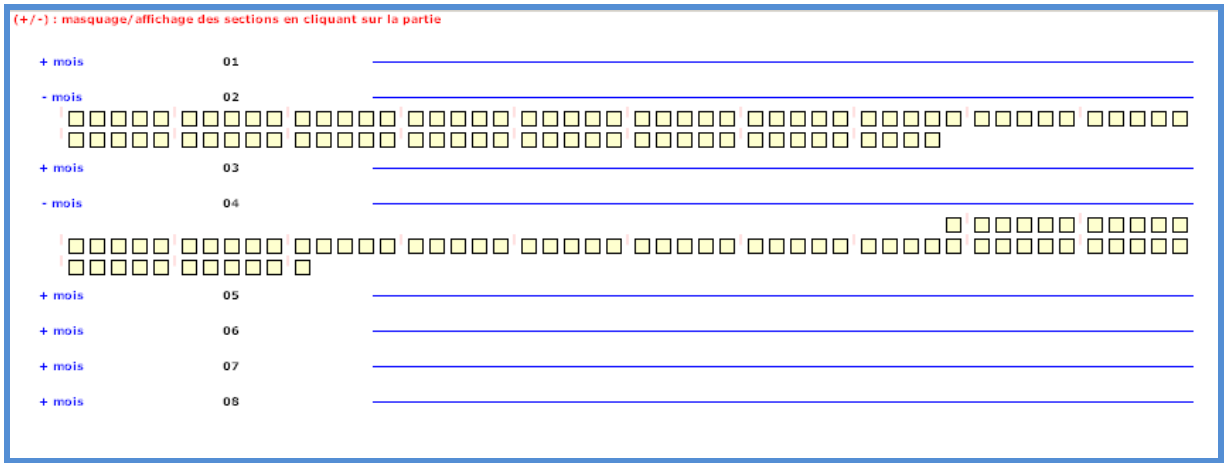


Figure 33 : *Le Trameur* Onglet SECTION, masquage de sections (2)

Pour inhiber tous les masquages, il faut réinitialiser la carte des sections via le bouton :

12 Accès à la *Trame* par la carte des sections : principales fonctionnalités

12.1 Edition de section

Une fois la carte construite, les sections sont éditables dans l'éditeur de l'onglet **SECTION**

- Soit en cliquant sur la section visée
- Soit en activant la combinaison *shift+clic* : cette solution permet de maintenir la sélection de la section visée pour des calculs ultérieurs (*cf infra*)

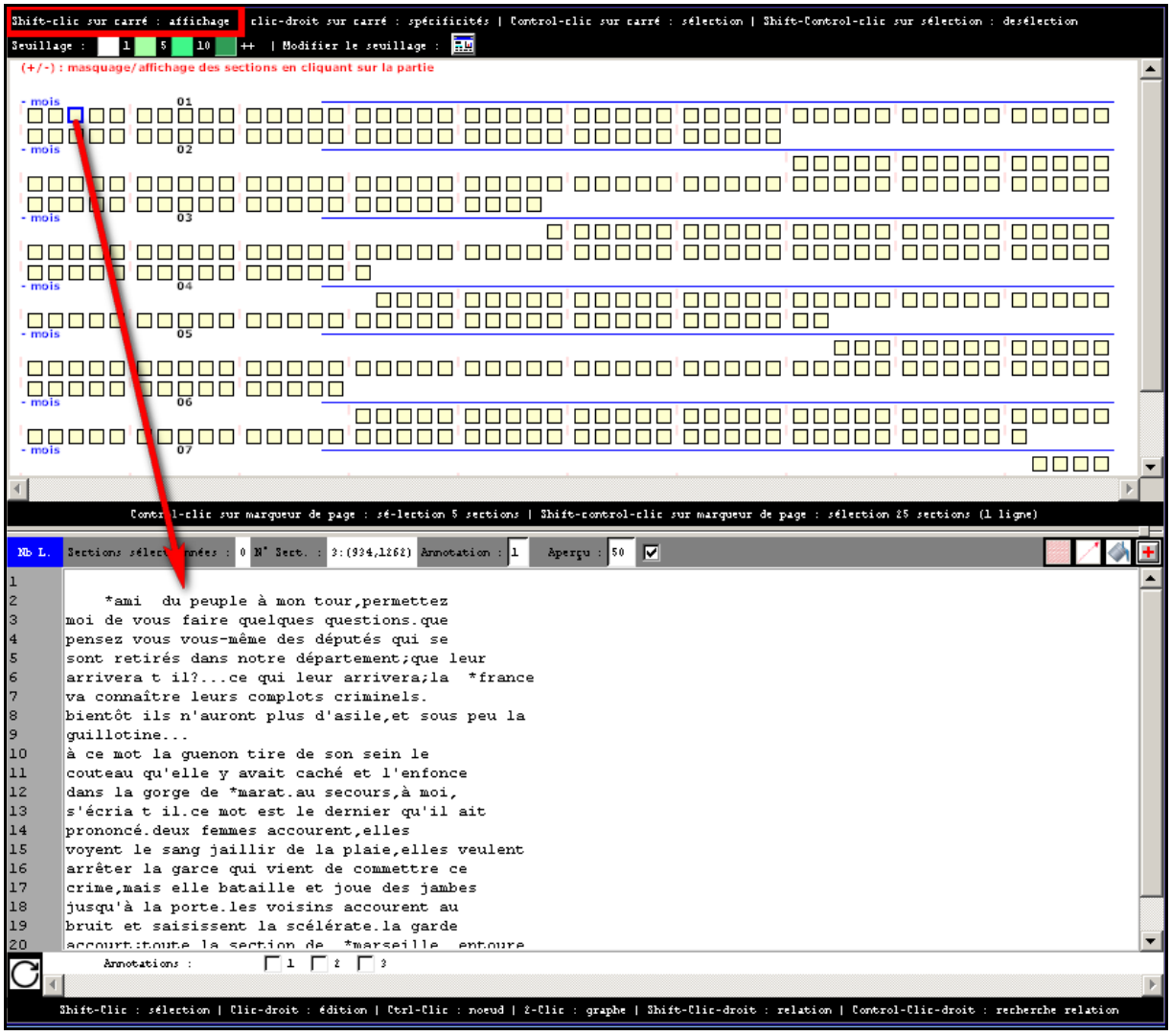
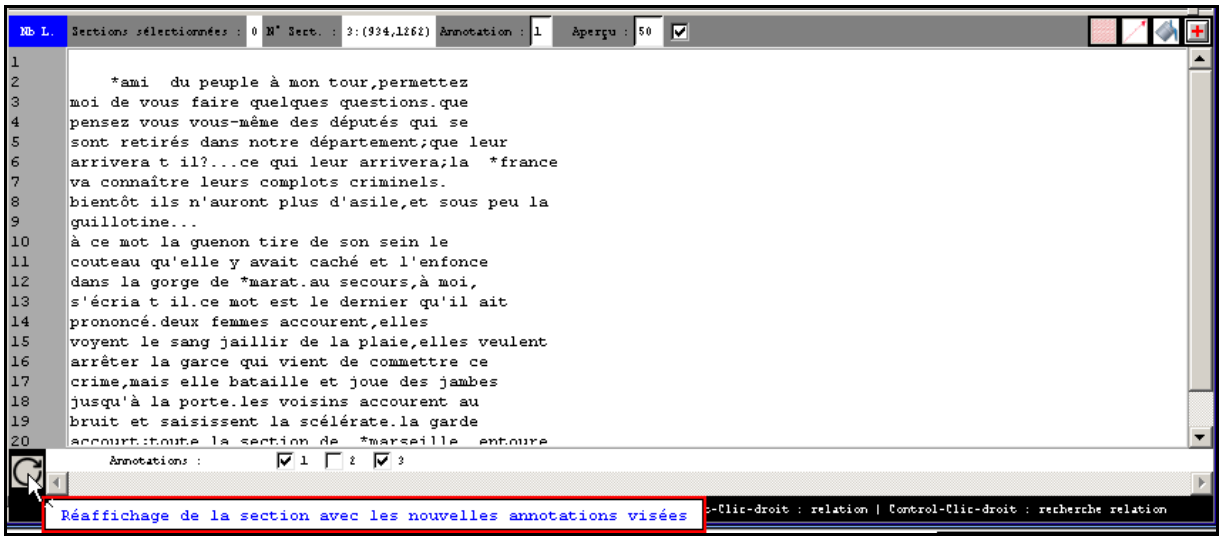

Shift-clic sur carré : affichage
clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : desélection
Seuillage :
Modifier le seuillage :
(+/-) : masquage/affichage des sections en cliquant sur la partie
- mois
01
02
03
04
05
06
07
Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)
Nb L.
Annotation :
Aperçu :
*ami du peuple à mon tour,permettez
moi de vous faire quelques questions.que
pensez vous vous-même des députés qui se
sont retirés dans notre département;que leur
arrivera t il?...ce qui leur arrivera;la *france
va connaître leurs complots criminels.
bientôt ils n'auront plus d'asile,et sous peu la
guillotine...
à ce mot la guenon tire de son sein le
couteau qu'elle y avait caché et l'enfonce
dans la gorge de *marat.au secours,à moi,
s'écria t il.ce mot est le dernier qu'il ait
prononcé.deux femmes accourent,elles
voyent le sang jaillir de la plaie,elles veulent
arrêter la garce qui vient de commettre ce
crime,mais elle bataille et joue des jambes
jusqu'à la porte.les voisins accourent au
bruit et saisissent la scélérate.la garde
Annotations :
Shift-Clic : sélection | Clic-droit : édition | Ctrl-Clic : noeud | 2-Clic : graphe | Shift-Clic-droit : relation | Control-Clic-droit : recherche relation

Figure 34 : *Le Trameur* Onglet SECTION, édition de section

On peut ensuite recharger l'affichage du contenu de la section en sélectionnant les annotations à visualiser :



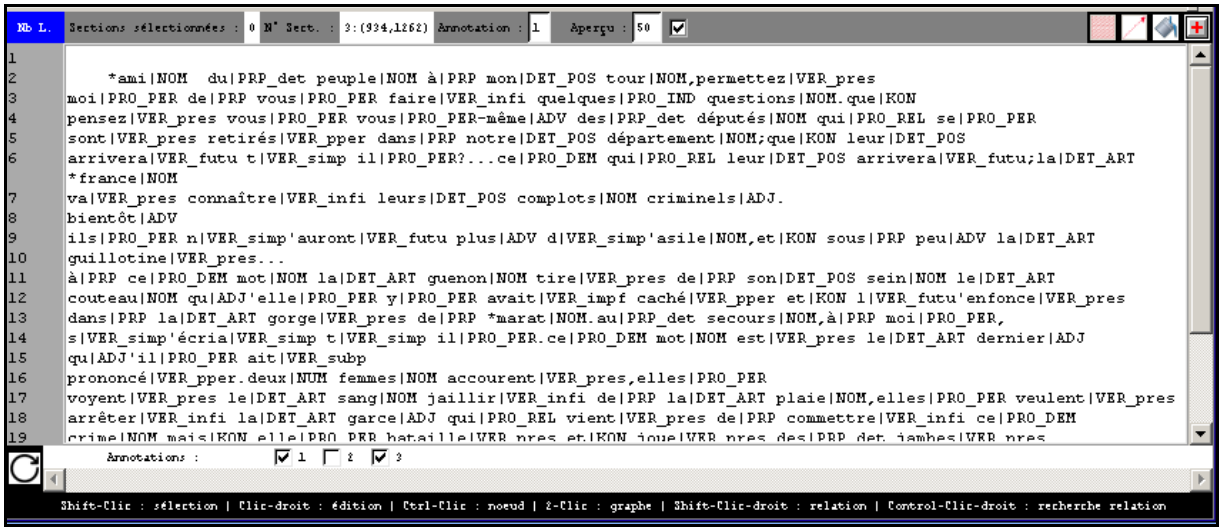


Figure 35 : *Le Trameur* Onglet SECTION, édition de section, sélection des annotations

12.2 Recherche de chaînes

On dispose dans l'onglet **SECTION** d'une zone de saisie permettant de lancer des recherches de chaînes : "*Ventilation de*". Pour rechercher une forme, il convient de l'insérer dans la zone de saisie puis de lancer la recherche en tapant la touche « Entrée ».

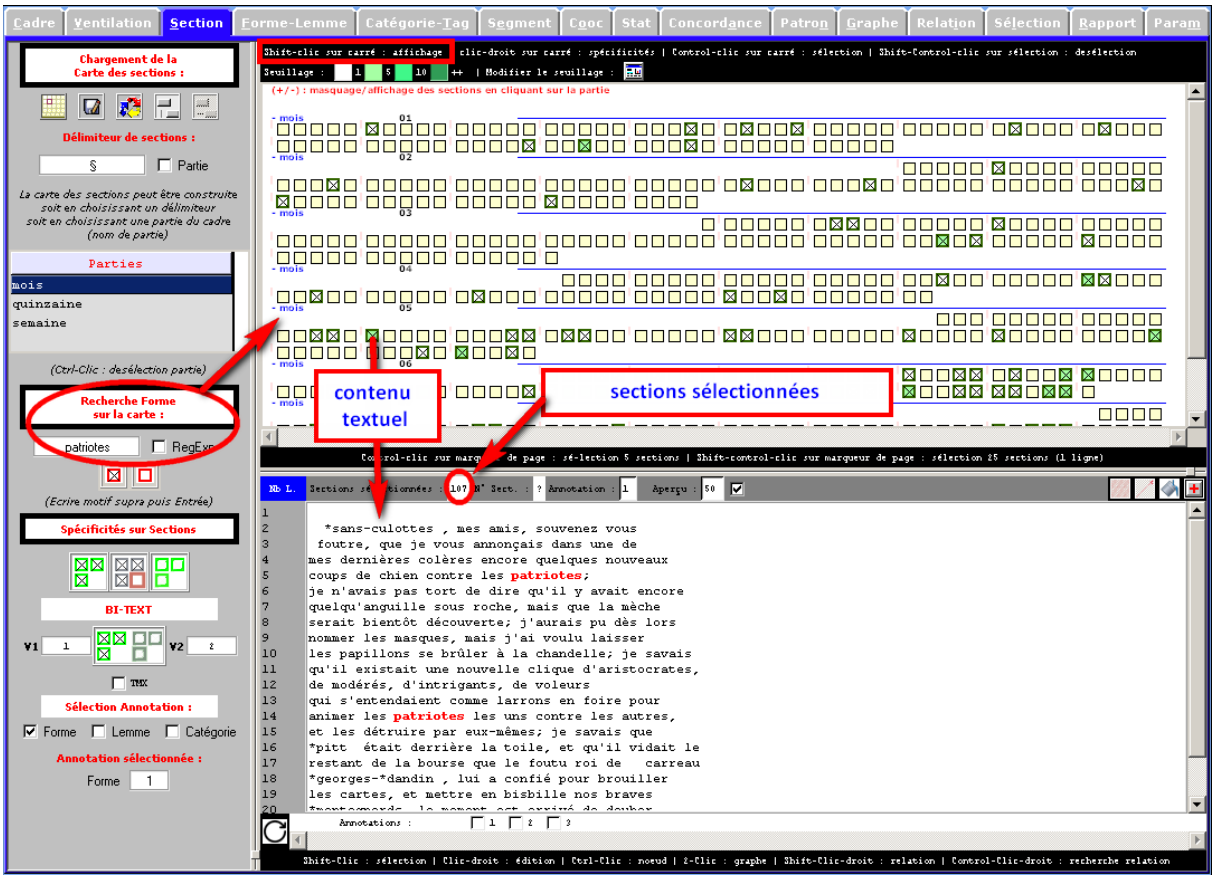


Figure 36 : *Le Trameur* Onglet SECTION, recherche de chaînes

Dans la figure précédente, la forme cherchée (dormeur) est présente dans la section représentée par le carré marqué d'une croix matérialisant la présence de la forme cherchée. La coloration verte de cette section signifie que celle-ci est désormais sélectionnée pour des opérations spécifiques sur cette section (calcul de spécificité, recherche de cooccurrents, *cf infra*).

Le contenu textuel est visible *via* le raccourci clavier *shift-clic* sur le carré associé à la section : ce raccourci permet de maintenir cette sélection sélectionnée contrairement au simple *clic* sur la section qui permet son affichage mais aussi sa désélection (*cf supra*).

On peut aussi lancer une recherche en sélectionnant une forme dans les dictionnaires disponibles (et *via* le raccourci clavier défini).

12.3 Recherche « incrémentale » de chaînes

Un mécanisme similaire à celui présenté ci-dessus sur l'éditeur de la ***Trame*** dans l'onglet **CADRE** est disponible sur l'onglet **SECTION**. Le raccourci clavier *Control-r* active ce mode de recherche permettant de définir *incrémentalement* un motif de recherche.

12.4 Visualisation des différentes annotations des items de la *Trame*

Un mécanisme similaire à celui présenté ci-dessus sur l'éditeur de la ***Trame*** dans l'onglet **CADRE** est disponible sur l'onglet **SECTION**.

L'éditeur de la ***Trame*** de l'onglet **SECTION** donne à voir les items de la ***Trame*** et les annotations associées.

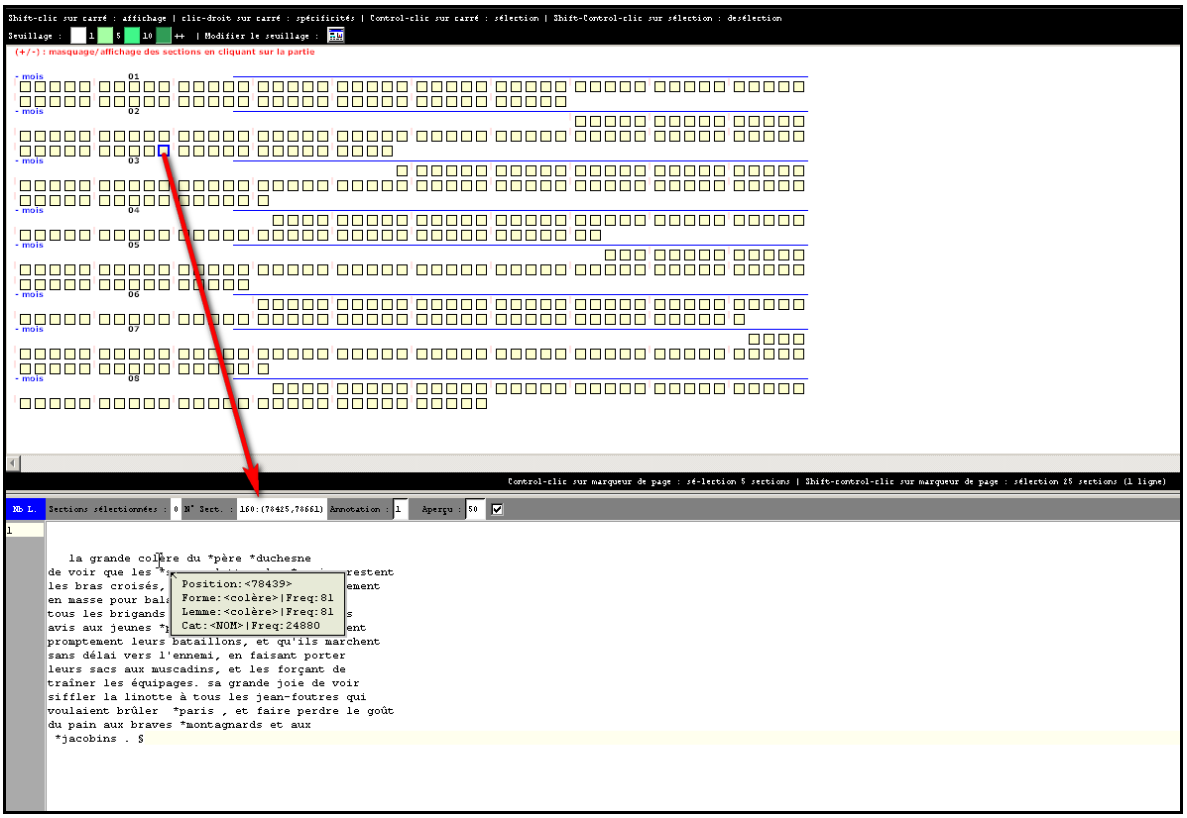

la grande colère du *père *duchesne
Position:<78439>
Forme:<colère>|Freq:81
Lemme:<colère>|Freq:81
Cat:<NOM>|Freq:24880
promptement leurs bataillons, et qu'ils marchent
sans délai vers l'ennemi, en faisant porter
leurs sacs aux muscadins, et les forçant de
traîner les équipages. sa grande joie de voir
siffler la linotte à tous les jean-foutres qui
voulaient brûler *paris , et faire perdre le goût
du pain aux braves *montagnards et aux
*jacobins . $

Figure 37 : *Le Trameur* Onglet SECTION, recherche de chaînes

12.5 Edition et modification des items de la *Trame*

De la même manière, on peut modifier les items de la ***Trame*** à travers l'éditeur de la ***Trame*** de l'onglet **SECTION**. Un *clic-droit* sur un item déclenche l'accès aux différentes annotations de l'item visé, toute modification sera enregistrée dans les dictionnaires associés.

12.6 Le vocabulaire spécifique d'une section

Un *clic droit* sur un carré déclenche le calcul du vocabulaire spécifique[10] de la section considérée.

Specificites du vocabulaire sur la partie choisie : <Section n°457 (227227,227507) | delimiteur de section : §>

Nombre d'occurrences du texte global = 142179
Nombre d'occurrences dans la partie visee = 128
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit :
C:/SERGEF~1/SF-WORK/07-DEVEL/ANR-TE~1/MKGRAP~1/textes/duchn/export/resultspecif-source-1417428081.txt

Item	Ind-Specif	Fq-Totale	Fq-Partie
on	4.9	856	6
plus	4.1	798	5
entraîner	4.0	1	1
sali	3.7	2	1
persécute	3.6	3	1
sachant	3.4	4	1
naufrage	3.4	4	1
égarer	3.4	4	1

Figure 38 : *Le Trameur* Onglet SECTION, les mots spécifiques d'une section

12.7 Le vocabulaire spécifique d'une sélection de sections de la carte

Il est possible de sélectionner une série de sections dans la carte des sections et de calculer le vocabulaire spécifique[11] de cette sélection. Pour sélectionner une section, il convient d'activer la combinaison de touches : *Control+clic*. Après avoir été sélectionné, le contour du carré associé à la section devient vert. Après avoir sélectionné une série de sections, on peut activer le calcul du vocabulaire spécifique de cette sélection *via* le bouton disponible dans l'onglet **SECTION**.

Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : desélection
Seuillage : 1 5 10 ++ | Modifier le seuillage :
(+/-) : masquage/affichage des sections en cliquant sur la partie
- mois 01
- mois 02
- mois 03
- mois 04
- mois 05
- mois 06
- mois 07
- mois 08

[10] Cf Partie « Vocabulaire spécificique » pour le paramétrage de ce calcul
[11] idem

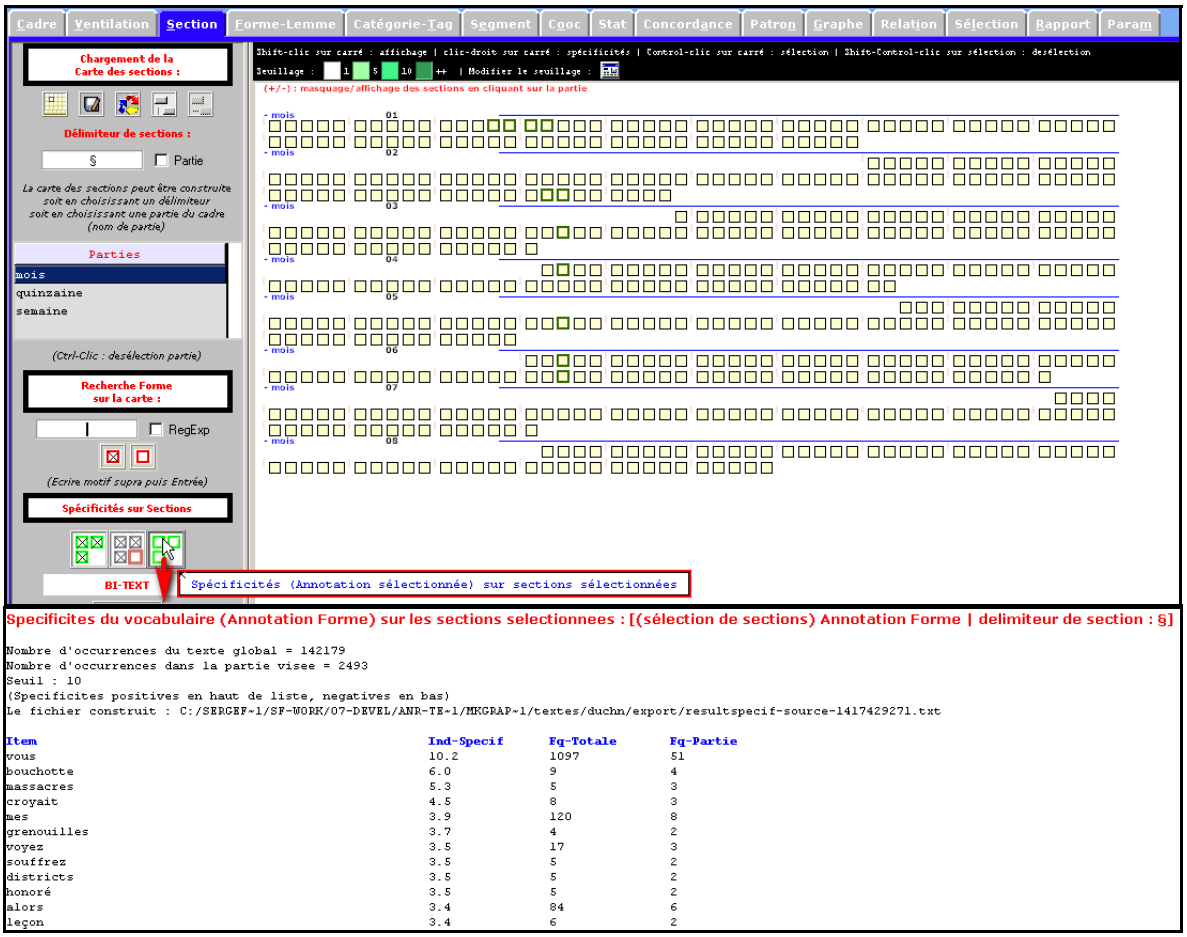
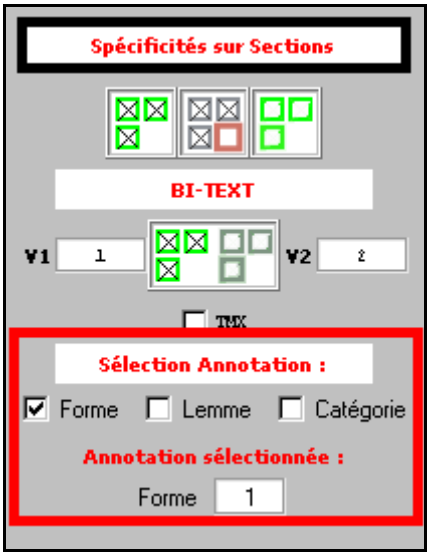


Figure 39 : *Le Trameur* Onglet SECTION, les mots spécifiques d'une sélection de sections

Ce calcul peut être fait sur tous les niveaux d'annotation de la base chargée :

Avant de lancer ce calcul, il est possible de sélectionner l'annotation utilisée pour réaliser le calcul. Pour cela il suffit de sélectionner l'annotation visée :

- soit en cochant l'annotation visée forme, lemme ou catégorie



- soit en donnant le numéro de l'annotation visée (puis entrée) dans la la zone de saisie

12.8 Le vocabulaire spécifique des sections de la carte contenant un motif : recherche de cooccurrents

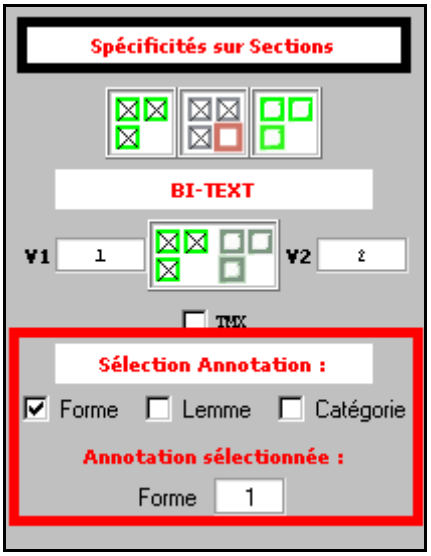
L'onglet **SECTION** dispose d'un bouton permettant de calculer les cooccurrents d'un item ou d'un motif donné.

⊠⊠ ⊠	recherche des cooccurrents d'un item (calcul des spécificités sur les sections contenant un motif donné)

Ce calcul peut être fait sur tous les niveaux d'annotation de la base chargée :

Avant de lancer ce calcul, il est possible de sélectionner l'annotation utilisée pour réaliser le calcul. Pour cela il suffit de sélectionner l'annotation visée :

- soit en cochant l'annotation visée forme, lemme ou catégorie

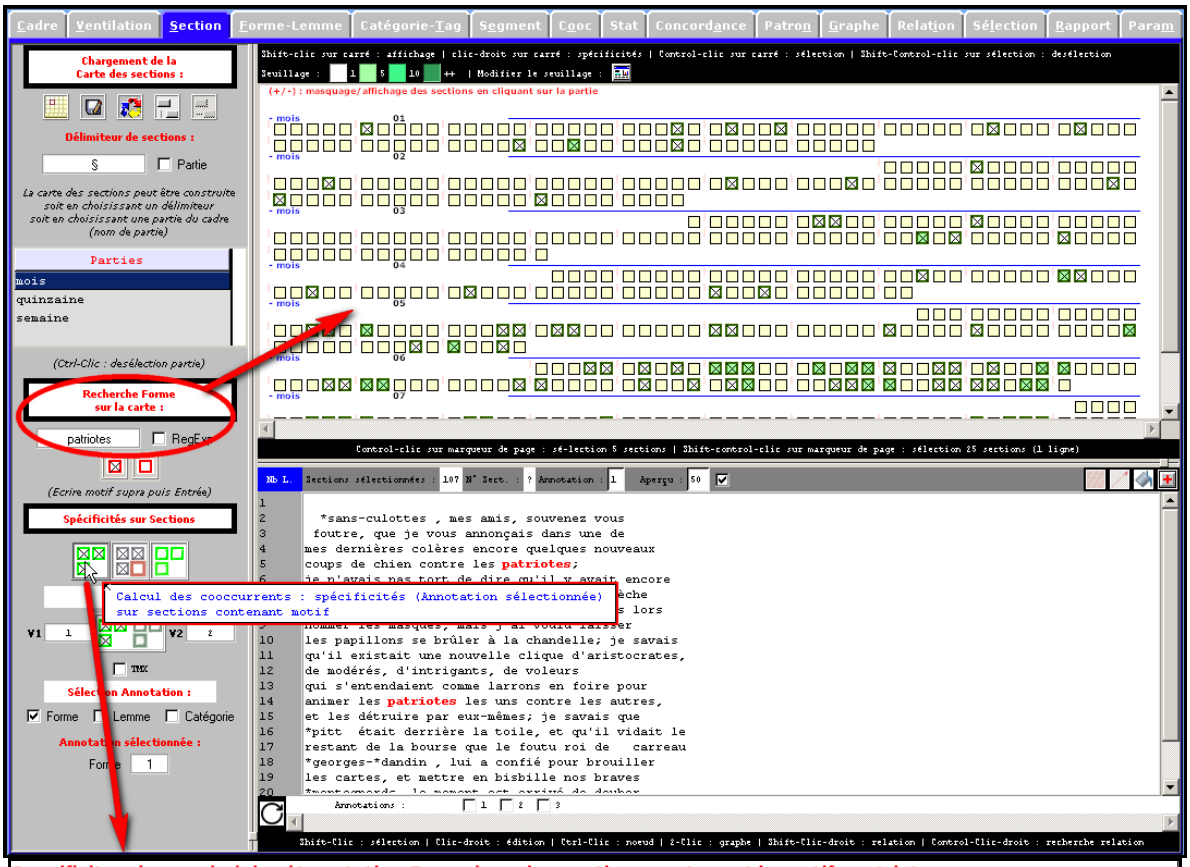


- soit en donnant le numéro de l'annotation visée (puis entrée) dans la la zone de saisie

Recherche d'un motif sur la carte puis recherche de cooccurrence :

A partir de la carte des sections, on peut rechercher un item ou un motif. On obtient une série de carré marqué par une croix indiquant la présence de l'item ou du motif. On peut ensuite lancer un calcul mettant au jour le vocabulaire spécifique de l'ensemble de ces sections. On obtient le vocabulaire spécifique[12] de l'ensemble des sections contenant l'item ou le motif initial : liste d'items qui portent soit un indice de spécificité positif soit un indice de spécificité négatif. Dans le premier cas, on aboutit en gros à une liste des cooccurrents de l'item ou du motif cherché, dans le second cas on obtient des items qui n'apparaissent pas ou peu avec l'item ou le motif cherché.

[12] *Cf* Partie « Vocabulaire spécificique » pour le paramétrage de ce calcul



Specificites du vocabulaire (Annotation Forme) sur les sections contenant le motif : patriotes

Nombre d'occurrences du texte global = 142179
Nombre d'occurrences dans la partie visee = 29871
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit :
C:/SERGEF~1/SF-WORK/07-DEVEL/ANR-TE~1/MKGRAP~1/textes/duchn/export/resultspecif-source-1417428336.txt

Item	Ind-Specif	Fq-Totale	Fq-Partie
patriotes	**	152	152
les	15.8	4749	1218
meilleurs	10.5	40	27
contre	8.4	298	103
chat	8.3	55	30
ont	8.2	669	199
brouiller	6.1	21	14
aristocrates	6.1	133	50
brissotins	5.8	62	28
vincent	5.7	15	11
jacobins	5.4	58	26
chien	5.4	101	39
ils	5.2	1187	305
cartes	5.2	32	17
fonctions	5.1	6	6
calomnie	5.0	12	9
accuse	5.0	12	9
feuillants	5.0	22	13
dis	4.8	34	17
purs	4.6	13	9
lyon	4.6	44	20
marseille	4.6	35	17
crête	4.5	11	8
tantôt	4.4	9	7
injures	4.4	9	7
confondus	4.3	7	6
accusent	4.2	14	9
faux	4.2	25	13
cordeliers	4.1	17	10
girondins	4.1	17	10
sauvé	4.0	23	12
midi	4.0	23	12
alors	3.9	84	30
intrigants	3.9	73	27
assassiner	3.9	10	7
diviser	3.7	16	9
énarmé	3.7	4	4

Figure 40 : *Le Trameur* Onglet SECTION, items spécifiques des sections contenant un motif (cooccurrents)

12.9 Le vocabulaire spécifique des sections de la carte ne contenant pas un motif

Un calcul de spécificité peut aussi être lancé sur les sections ne contenant pas le motif cherché dans la carte des sections :

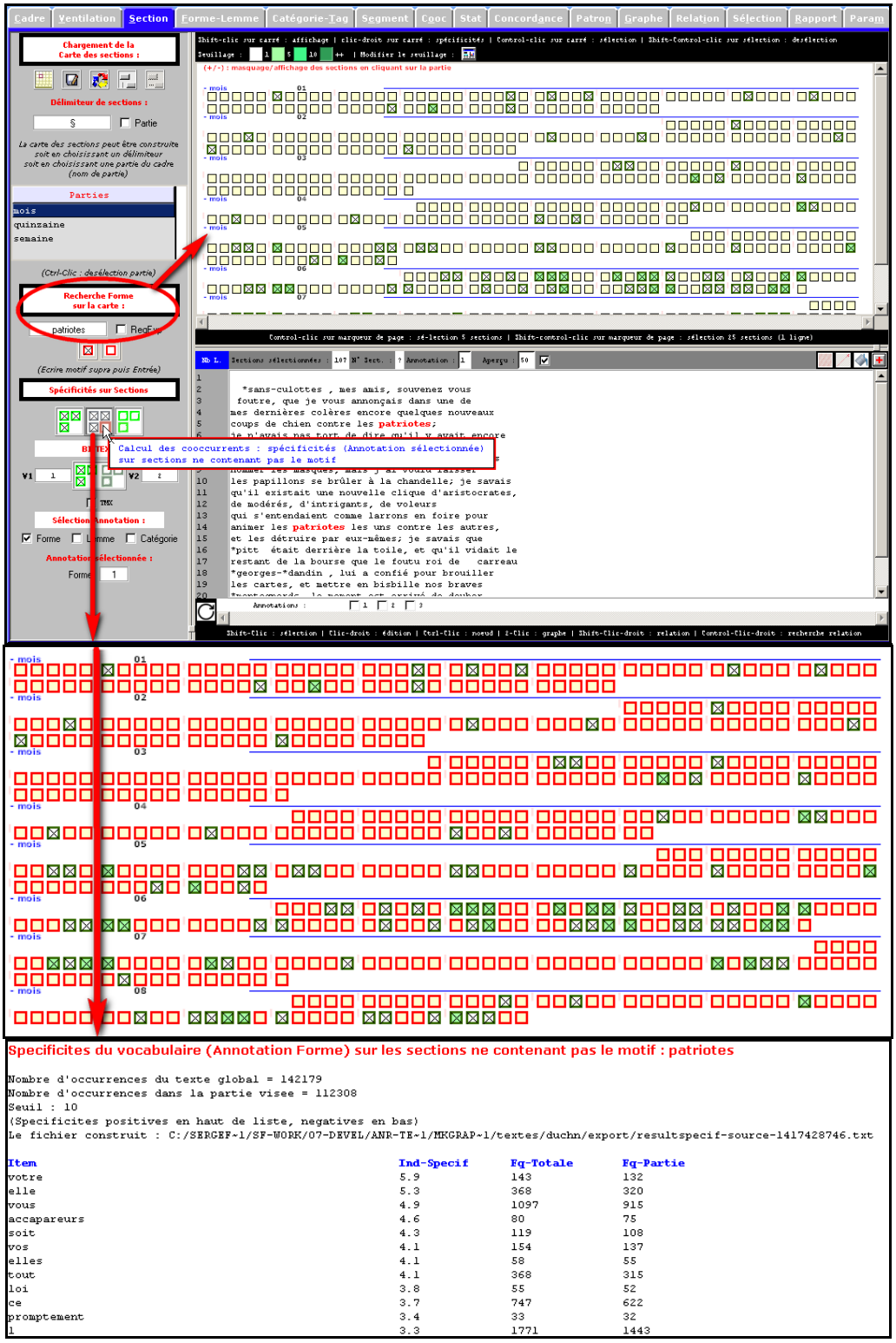

Cadre
Ventilation
Section
Forme-Lemme
Catégorie-Tag
Segment
Cooc
Stat
Concordance
Patron
Graphe
Relation
Sélection
Rapport
Param
Chargement de la Carte des sections :
Délimiteur de sections :
Partie
Parties
mois
quinzaine
semaine
Recherche Forme sur la carte :
patriotes
Spécificités sur Sections
Sélection Annotation :
Forme
Lemme
Catégorie
Annotation sélectionnée :
Calcul des cooccurrents : spécificités (Annotation sélectionnée) sur sections ne contenant pas le motif
"sans-culottes , mes amis, souvenez vous
foutre, que je vous annonçais dans une de
mes dernières colères encore quelques nouveaux
coups de chien contre les patriotes;
les papillons se brûler à la chandelle; je savais
qu'il existait une nouvelle clique d'aristocrates,
de modérés, d'intrigants, de voleurs
qui s'entendaient comme larrons en foire pour
animer les patriotes les uns contre les autres,
et les détruire par eux-mêmes; je savais que
*pitt était derrière la toile, et qu'il vidait le
restant de la bourse que le foutu roi de carreau
*georges-*dandin , lui a confié pour brouiller
les cartes, et mettre en bisbille nos braves
Specificites du vocabulaire (Annotation Forme) sur les sections ne contenant pas le motif : patriotes
Nombre d'occurrences du texte global = 142179
Nombre d'occurrences dans la partie visee = 112308
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit : C:/SERGEF~1/SF-WORK/07-DEVEL/ANR-TE~1/MKGRAP~1/textes/duchn/export/resultspecif-source-1417428746.txt
Item Ind-Specif Fq-Totale Fq-Partie
votre 5.9 143 132
elle 5.3 368 320
vous 4.9 1097 915
accapareurs 4.6 80 75
soit 4.3 119 108
vos 4.1 154 137
elles 4.1 58 55
tout 4.1 368 315
loi 3.8 55 52
ce 3.7 747 622
promptement 3.4 33 32
1 3.3 1771 1443

Figure 41 : *Le Trameur* Onglet SECTION, items spécifiques des sections ne contenant pas un motif

Là encore, ce calcul peut être fait sur tous les niveaux d'annotation de la base chargée : *cf supra*.

12.10 Fonctionnalités complémentaires sur la carte des sections

Control-Clic sur le trait vertical entre 2 séries de 5 sections : sélection des 5 sections à sa droite

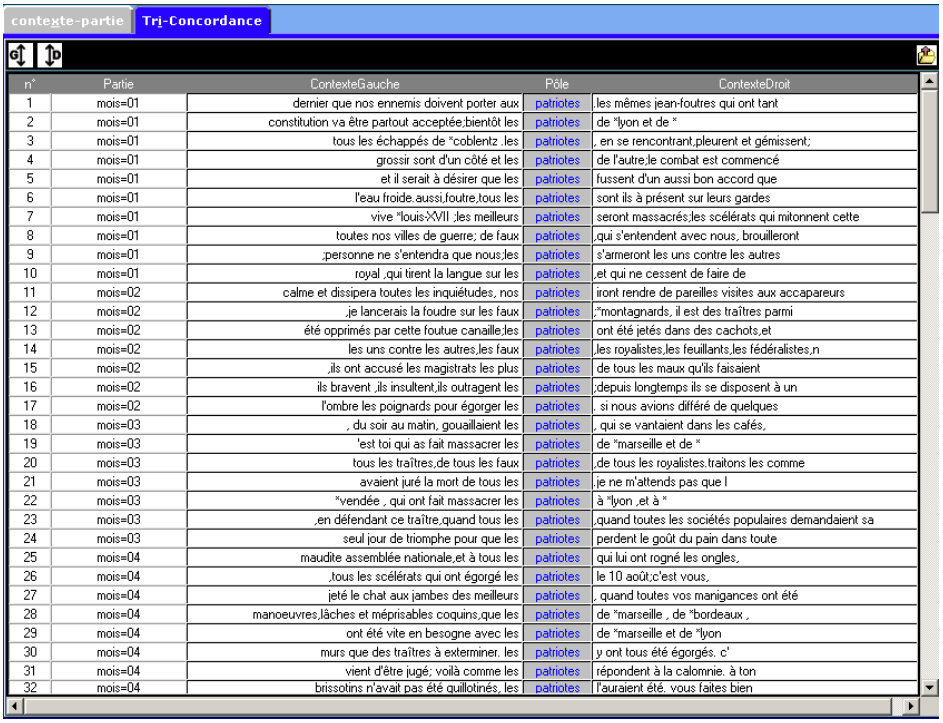
Shift–Control-Clic sur ce même trait vertical : sélection des 25 sections à sa droite (une ligne dans la carte des sections)

Shift-Control-Clic sur un carré dans la carte des sections : désélection de la section visée.

13 Concordances

L'onglet **CONCORDANCE** permet d'afficher les concordances des différentes annotations disponibles sur chacun des items de la ***Trame*** : forme, lemme, catégorie, annotations complémentaires. Les concordances produites sont disponibles dans 2 sous-onglets distincts : un sous-onglet avec édition des items de la concordance, un sous-onglet permettant de trier les contextes de la concordance.

contexte-partie
Tri-Concordance
Concordance de <sélection Positions>
mois=01
dernier que nos ennemis doivent porter aux patriotes.les mêmes jean-foutres qui ont tant
constitution va être partout acceptée;bientôt les patriotes de *lyon et de *
tous les échappés de *coblentz .les patriotes, en se rencontrant,pleurent et gémissent;
grossir sont d'un côté et les patriotes de l'autre;le combat est commencé
et il serait à désirer que les patriotes fussent d'un aussi bon accord que
l'eau froide.aussi,foutre,tous les patriotes sont ils à présent sur leurs gardes
vive *louis-XVII ;les meilleurs patriotes seront massacrés;les scélérats qui mitonnent cette
toutes nos villes de guerre; de faux patriotes,qui s'entendent avec nous, brouilleront
;personne ne s'entendra que nous;les patriotes s'armeront les uns contre les autres
royal ,qui tirent la langue sur les patriotes,et qui ne cessent de faire de
mois=02
calme et dissipera toutes les inquiétudes, nos patriotes iront rendre de pareilles visites aux accapareurs
,je lancerais la foudre sur les faux patriotes;*montagnards, il est des traîtres parmi
été opprimés par cette foutue canaille;les patriotes ont été jetés dans des cachots,et
les uns contre les autres,les faux patriotes,les royalistes,les feuillants,les fédéralistes,n
,ils ont accusé les magistrats les plus patriotes de tous les maux qu'ils faisaient
ils bravent ,ils insultent,ils outragent les patriotes;depuis longtemps ils se disposent à un
l'ombre les poignards pour égorger les patriotes. si nous avions différé de quelques
mois=03
, du soir au matin, gouaillaient les patriotes, qui se vantaient dans les cafés,
'est toi qui as fait massacrer les patriotes de *marseille et de *
tous les traîtres,de tous les faux patriotes,de tous les royalistes.traitons les comme
avaient juré la mort de tous les patriotes.je ne m'attends pas que l
*vendée , qui ont fait massacrer les patriotes à *lyon ,et à *
,en défendant ce traître,quand tous les patriotes,quand toutes les sociétés populaires demandaient sa
seul jour de triomphe pour que les patriotes perdent le goût du pain dans toute
mois=04
maudite assemblée nationale,et à tous les patriotes qui lui ont rogné les ongles,
,tous les scélérats qui ont égorgé les patriotes le 10 août;c'est vous,
jeté le chat aux jambes des meilleurs patriotes, quand toutes vos manigances ont été
manoeuvres,lâches et méprisables coquins,que les patriotes de *marseille , de *bordeaux ,
ont été vite en besogne avec les patriotes de *marseille et de *lyon
murs que des traîtres à exterminer. les patriotes y ont tous été égorgés. c'
vient d'être jugé; voilà comme les patriotes répondent à la calomnie. à ton
brissotins n'avait pas été guillotinés, les patriotes l'auraient été. vous faites bien
mois=05
*samson à s'ensevelir avec les patriotes sous les ruines du temple. ils
la patte graissée pour assassiner les meilleurs patriotes. $ il y a encore



n°	Partie	ContexteGauche	Pôle	ContexteDroit
1	mois=01	dernier que nos ennemis doivent porter aux	patriotes	.les mêmes jean-foutres qui ont tant
2	mois=01	constitution va être partout acceptée;bientôt les	patriotes	de *lyon et de *
3	mois=01	tous les échappés de *coblentz .les	patriotes	, en se rencontrant,pleurent et gémissent;
4	mois=01	grossir sont d'un côté et les	patriotes	de l'autre;le combat est commencé
5	mois=01	et il serait à désirer que les	patriotes	fussent d'un aussi bon accord que
6	mois=01	l'eau froide.aussi,foutre,tous les	patriotes	sont ils à présent sur leurs gardes
7	mois=01	vive *louis-XVII ;les meilleurs	patriotes	seront massacrés;les scélérats qui mitonnent cette
8	mois=01	toutes nos villes de guerre; de faux	patriotes	,qui s'entendent avec nous, brouilleront
9	mois=01	;personne ne s'entendra que nous;les	patriotes	s'armeront les uns contre les autres
10	mois=01	royal ,qui tirent la langue sur les	patriotes	,et qui ne cessent de faire de
11	mois=02	calme et dissipera toutes les inquiétudes, nos	patriotes	iront rendre de pareilles visites aux accapareurs
12	mois=02	,je lancerais la foudre sur les faux	patriotes	;*montagnards, il est des traîtres parmi
13	mois=02	été opprimés par cette foutue canaille;les	patriotes	ont été jetés dans des cachots,et
14	mois=02	les uns contre les autres,les faux	patriotes	,les royalistes,les feuillants,les fédéralistes,n
15	mois=02	,ils ont accusé les magistrats les plus	patriotes	de tous les maux qu'ils faisaient
16	mois=02	ils bravent ,ils insultent,ils outragent les	patriotes	;depuis longtemps ils se disposent à un
17	mois=02	l'ombre les poignards pour égorger les	patriotes	. si nous avions différé de quelques
18	mois=03	, du soir au matin, gouaillaient les	patriotes	, qui se vantaient dans les cafés,
19	mois=03	'est toi qui as fait massacrer les	patriotes	de *marseille et de *
20	mois=03	tous les traîtres,de tous les faux	patriotes	,de tous les royalistes.traitons les comme
21	mois=03	avaient juré la mort de tous les	patriotes	,je ne m'attends pas que l
22	mois=03	*vendée , qui ont fait massacrer les	patriotes	à *lyon ,et à *
23	mois=03	,en défendant ce traître,quand tous les	patriotes	,quand toutes les sociétés populaires demandaient sa
24	mois=03	seul jour de triomphe pour que les	patriotes	perdent le goût du pain dans toute
25	mois=04	maudite assemblée nationale,et à tous les	patriotes	qui lui ont rogné les ongles,
26	mois=04	,tous les scélérats qui ont égorgé les	patriotes	le 10 août;c'est vous,
27	mois=04	jeté le chat aux jambes des meilleurs	patriotes	, quand toutes vos manigances ont été
28	mois=04	manoeuvres,lâches et méprisables coquins,que les	patriotes	de *marseille , de *bordeaux ,
29	mois=04	ont été vite en besogne avec les	patriotes	de *marseille et de *lyon
30	mois=04	murs que des traîtres à exterminer. les	patriotes	y ont tous été égorgés. c'
31	mois=04	vient d'être jugé; voilà comme les	patriotes	répondent à la calomnie. à ton
32	mois=04	brissotins n'avait pas été guillotinés, les	patriotes	l'auraient été. vous faites bien

Figure 42 : *Le Trameur* Onglet CONCORDANCE, affichage édition + affichage tri

13.1 Modes de déclenchement d'une concordance

Ces concordances peuvent être réalisées :

- à partir des dictionnaires, d'une liste de segments répétés, d'une liste de séquences de termes etc. en sélectionnant les items visés dans la liste visée et en activant la requête de concordance à partir du bouton disponible dans l'onglet associé,
- ou en renseignant directement les zones de saisie idoines décrites dans les figures qui suivent.

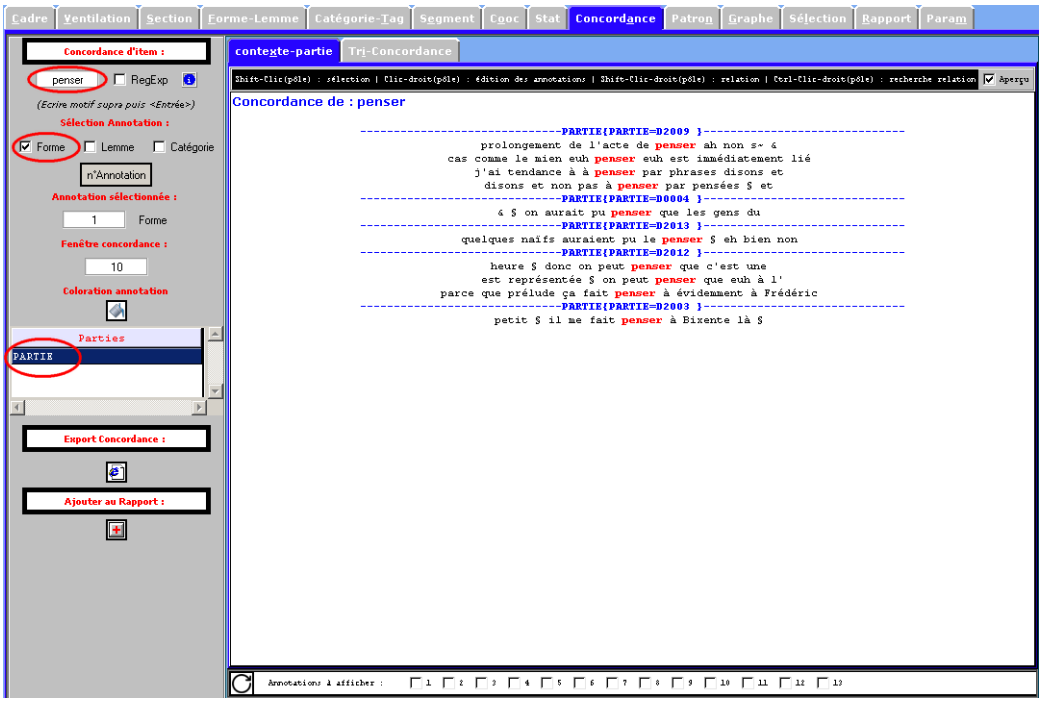


Figure 43 : ***Le Trameur*** **Onglet CONCORDANCE** (forme)

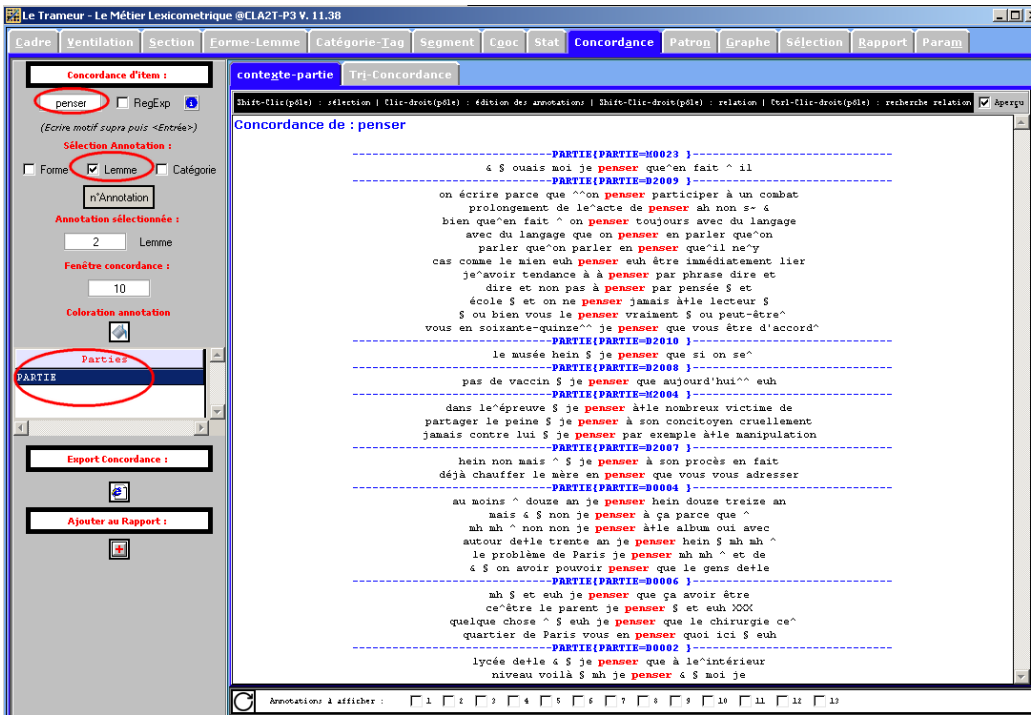

Le Trameur - Le Métier Lexicometrique @CLA2T-P3 V. 11.38
Cadre
Ventilation
Section
Forme-Lemme
Catégorie-Tag
Segment
Cooc
Stat
Concordance
Patron
Graphe
Sélection
Rapport
Param
Concordance d'item :
penser
RegExp
(Ecrire motif supra puis <Entrée>)
Sélection Annotation :
Forme
Lemme
Catégorie
n°Annotation
Annotation sélectionnée :
2
Lemme
Fenêtre concordance :
10
Coloration annotation
Parties
PARTIE
Export Concordance :
Ajouter au Rapport :
contexte-partie
Tri-Concordance
Concordance de : penser
PARTIE{PARTIE=M0023 }
& $ ouais moi je penser que^en fait ^ il
PARTIE{PARTIE=D2009 }
on écrire parce que ^^on penser participer à un combat
prolongement de le^acte de penser ah non s~ &
bien que^en fait ^ on penser toujours avec du langage
avec du langage que on penser en parler que^on
parler que^on parler en penser que^il ne~y
cas comme le mien euh penser euh être immédiatement lier
je^avoir tendance à à penser par phrase dire et
dire et non pas à penser par pensée $ et
école $ et on ne penser jamais à+le lecteur $
$ ou bien vous le penser vraiment $ ou peut-être^
vous en soixante-quinze^^ je penser que vous être d'accord^
PARTIE{PARTIE=D2010 }
le musée hein $ je penser que si on se^
PARTIE{PARTIE=D2008 }
pas de vaccin $ je penser que aujourd'hui^^ euh
PARTIE{PARTIE=M2004 }
dans le^épreuve $ je penser à+le nombreux victime de
partager le peine $ je penser à son concitoyen cruellement
jamais contre lui $ je penser par exemple à+le manipulation
PARTIE{PARTIE=D2007 }
hein non mais ^ $ je penser à son procès en fait
déjà chauffer le mère en penser que vous vous adresser
PARTIE{PARTIE=D0004 }
au moins ^ douze an je penser hein douze treize an
mais & $ non je penser à ça parce que ^
mh mh ^ non non je penser à+le album oui avec
autour de+le trente an je penser hein $ mh mh ^
le problème de Paris je penser mh mh ^ et de
& $ on avoir pouvoir penser que le gens de+le
PARTIE{PARTIE=D0006 }
mh $ et euh je penser que ça avoir être
ce^être le parent je penser $ et euh XXX
quelque chose ^ $ euh je penser que le chirurgie ce^
quartier de Paris vous en penser quoi ici $ euh
PARTIE{PARTIE=D0002 }
lycée de+le & $ je penser que à le^intérieur
niveau voilà $ mh je penser & $ moi je
Annotations à afficher :

Figure 44 : ***Le Trameur*** **Onglet CONCORDANCE** (lemme)

Figure 45 : *Le Trameur* Onglet CONCORDANCE (catégorie)

13.2 Paramétrage d'une concordance

L'affichage de la concordance peut être paramétré de la manière suivante :

- choisir une ou plusieurs parties
- déterminer la longueur de la fenêtre graphique de la concordance (par défaut 10 formes à droite et 10 à gauche)
- choisir le type de forme à rechercher en cochant l'une des cases : Forme, Lemme, Catégorie (ou annotation importée *cf infra*)
- choisir l'item visé (une forme, un lemme, une valeur de catégorie ou une annotation importée)

Dans les exemples précédents, l'affichage demandé est celui d'une forme, d'un lemme puis d'une catégorie en tenant compte de la seule partition disponible.

On peut aussi activer la concordance en utilisant le mode *RegExp* : la recherche de la forme visée se fait en écrivant un motif exprimé sous la forme d'une expression régulière.

Une fois la concordance chargée, il est possible de la recharger en modifiant les annotations à afficher pour cette concordance. On trouve au bas de la fenêtre du concordancier une liste de cases à cocher, chacune d'elle est associée à une des annotations disponibles : en activant tel ou tel bouton, le rechargement de la concordance intègre l'affichage des annotations sélectionnées.

Si on considère cette première concordance (seule l'annotation n°1 (forme) est affichée) :

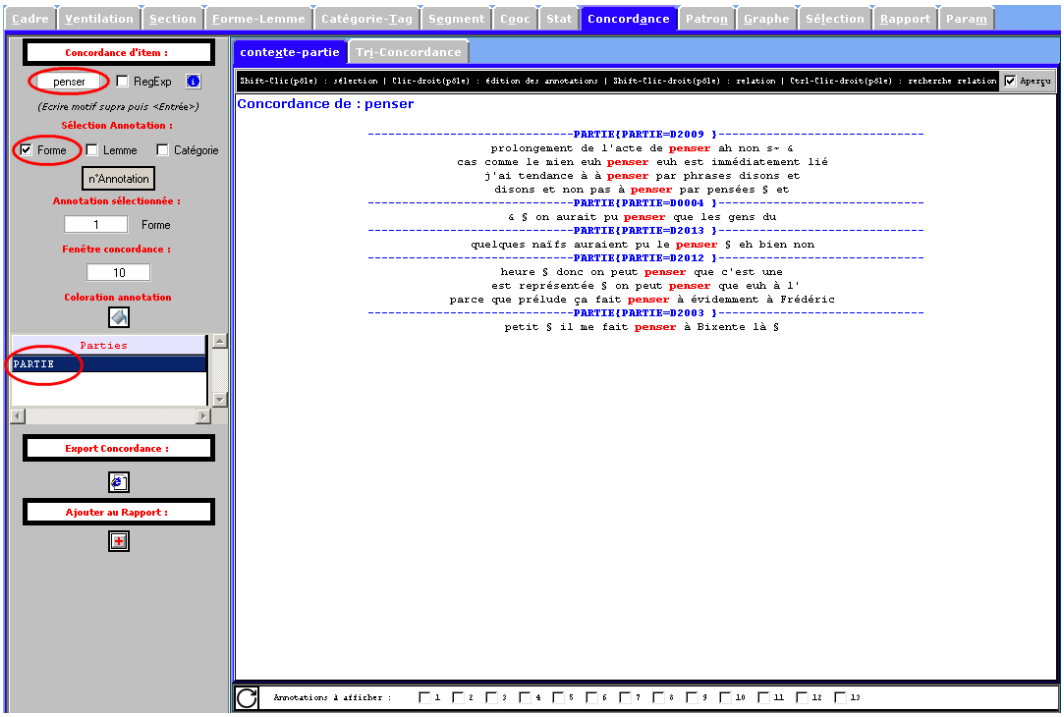


Figure 46 : *Le Trameur* Onglet CONCORDANCE (forme)

La sélection de nouvelles annotations et le rechargement de la concordance modifie l'affichage de cette concordance :

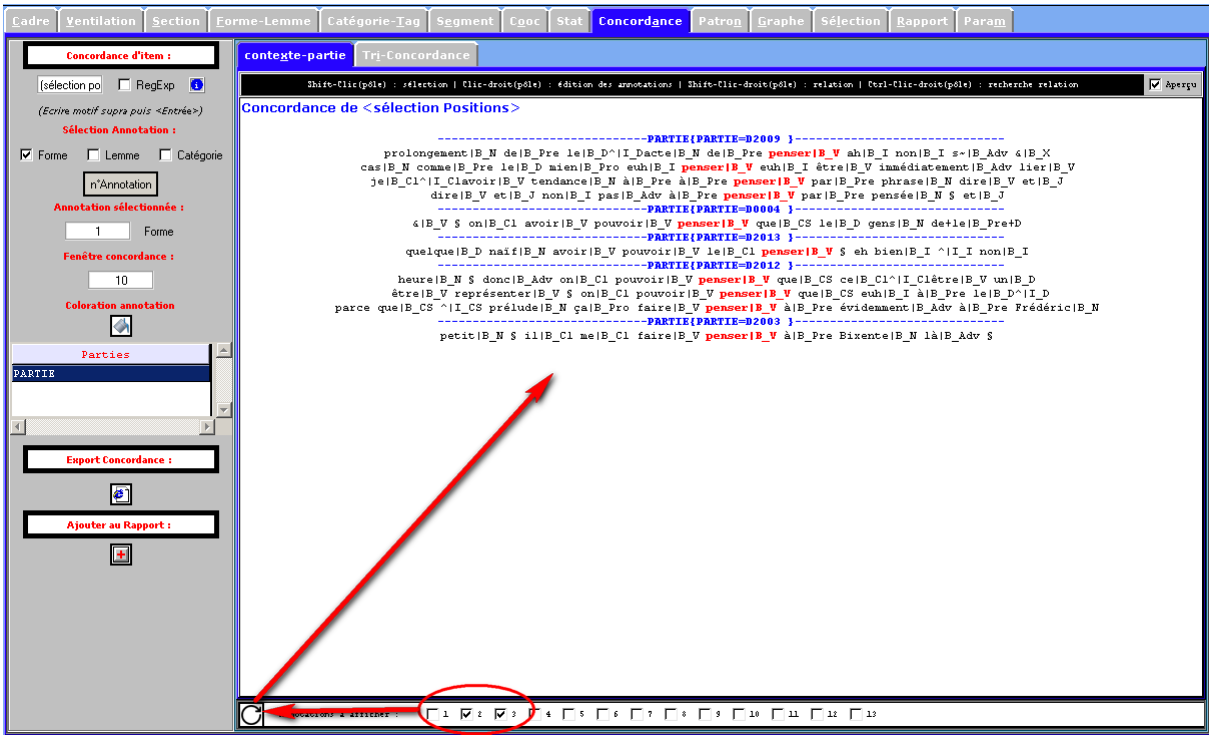


Figure 47 : *Le Trameur* Onglet CONCORDANCE (rechargement)

En activant le bouton , il est possible d'exporter une concordance au format HTML.

13.3 Editeur de concordance

Comme dans l'onglet **CADRE** et dans l'onglet **SECTION**, il est possible dans l'onglet **CONCORDANCE** de modifier les annotations associées à tous les items de la concordance affichée.

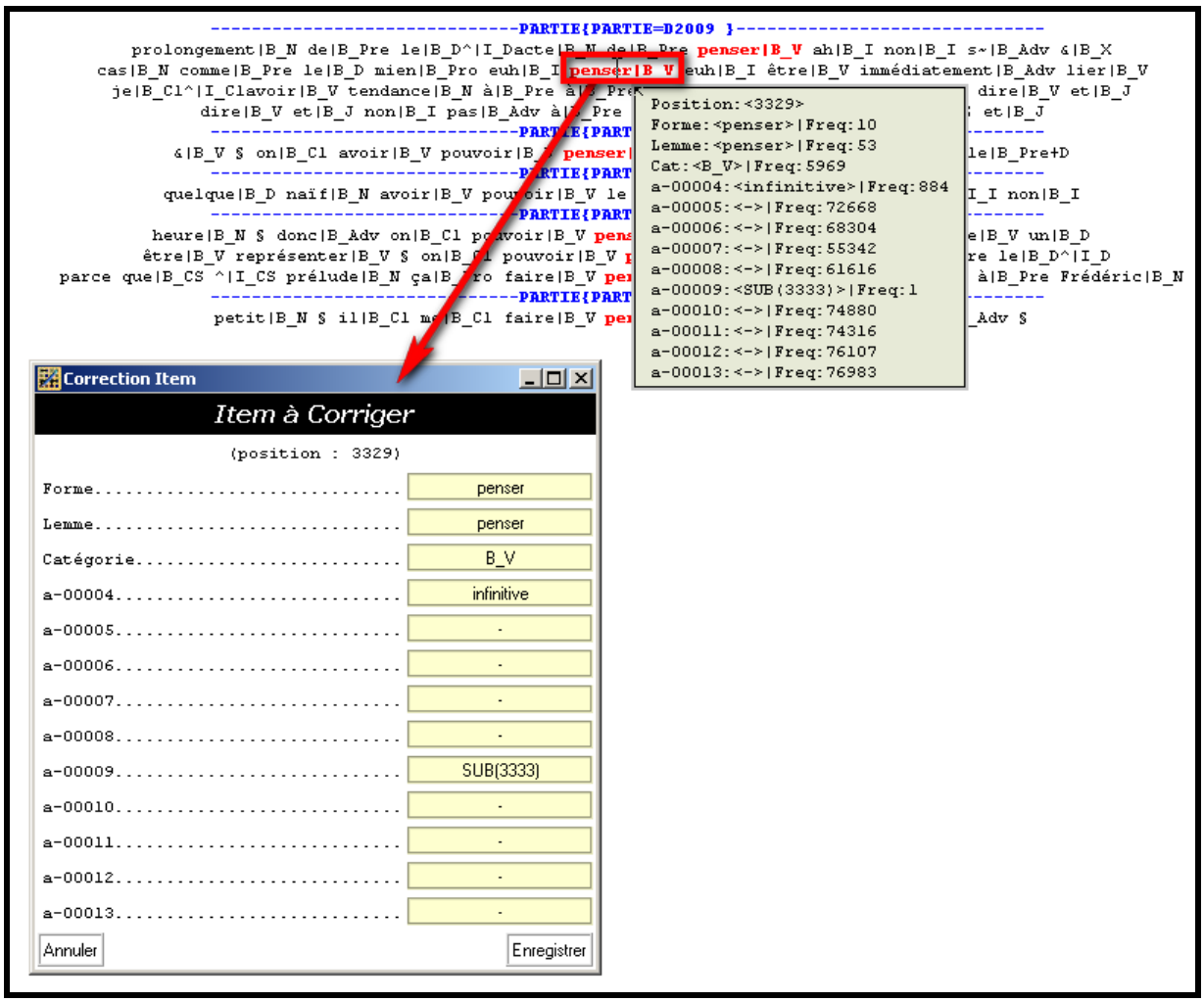


Figure 48 : *Le Trameur* Onglet CONCORDANCE, édition et modification d'un item pivot

Toute modification entraîne le rechargement des différentes ressources impactées par ces modifications.

14 Marquage des annotations de la *Trame*

14.1 Annotation « simple »

Le ***Trameur*** dispose de 3 éditeurs vus *supra* :

- **Editeur du *Cadre***
- **Editeur de la Carte des sections**
- **Editeur du concordancier**

Ces 3 éditeurs disposent d'une fonctionnalité permettant de colorer dans la zone d'édition les items possédant une annotation donnée. Cette fonctionnalité est activée par le bouton présent dans chacun des onglets contenant ces éditeurs.

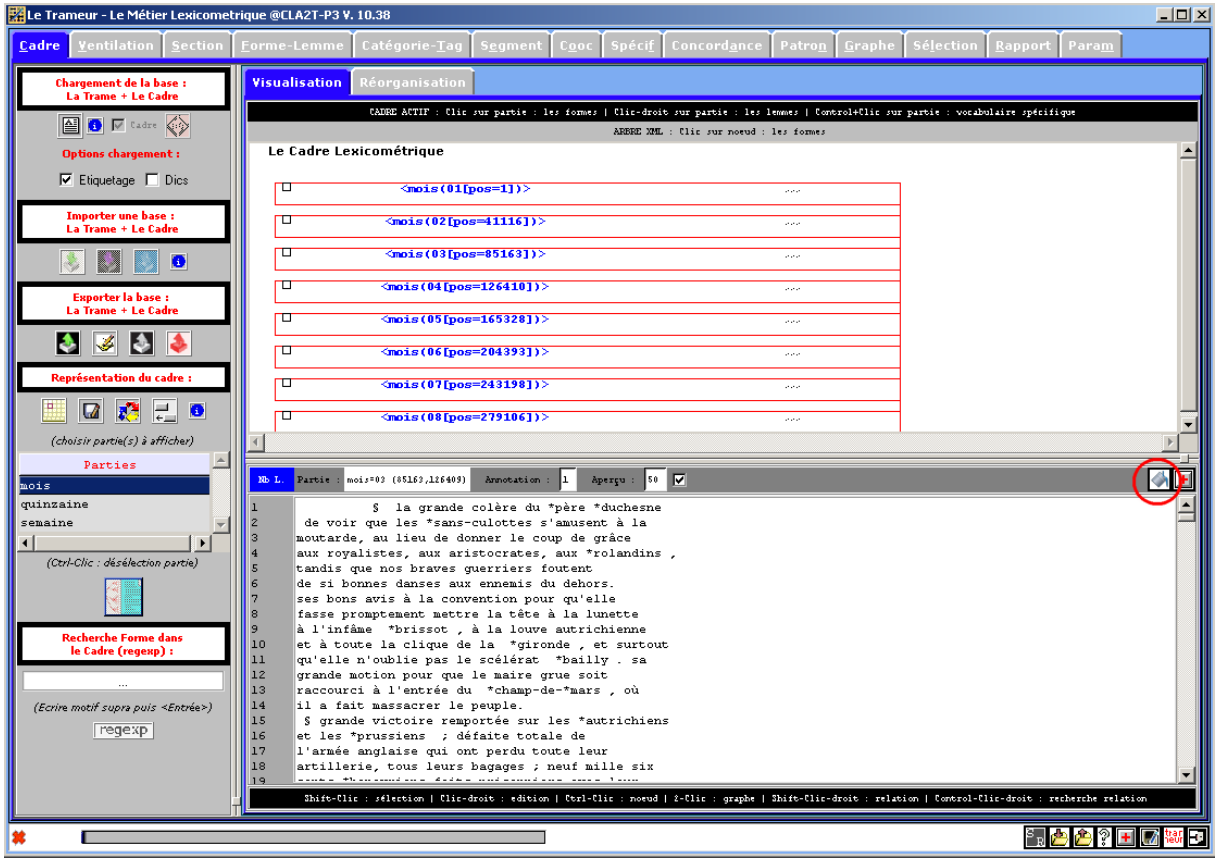


Figure 49 : *Le Trameur* Editeur CADRE

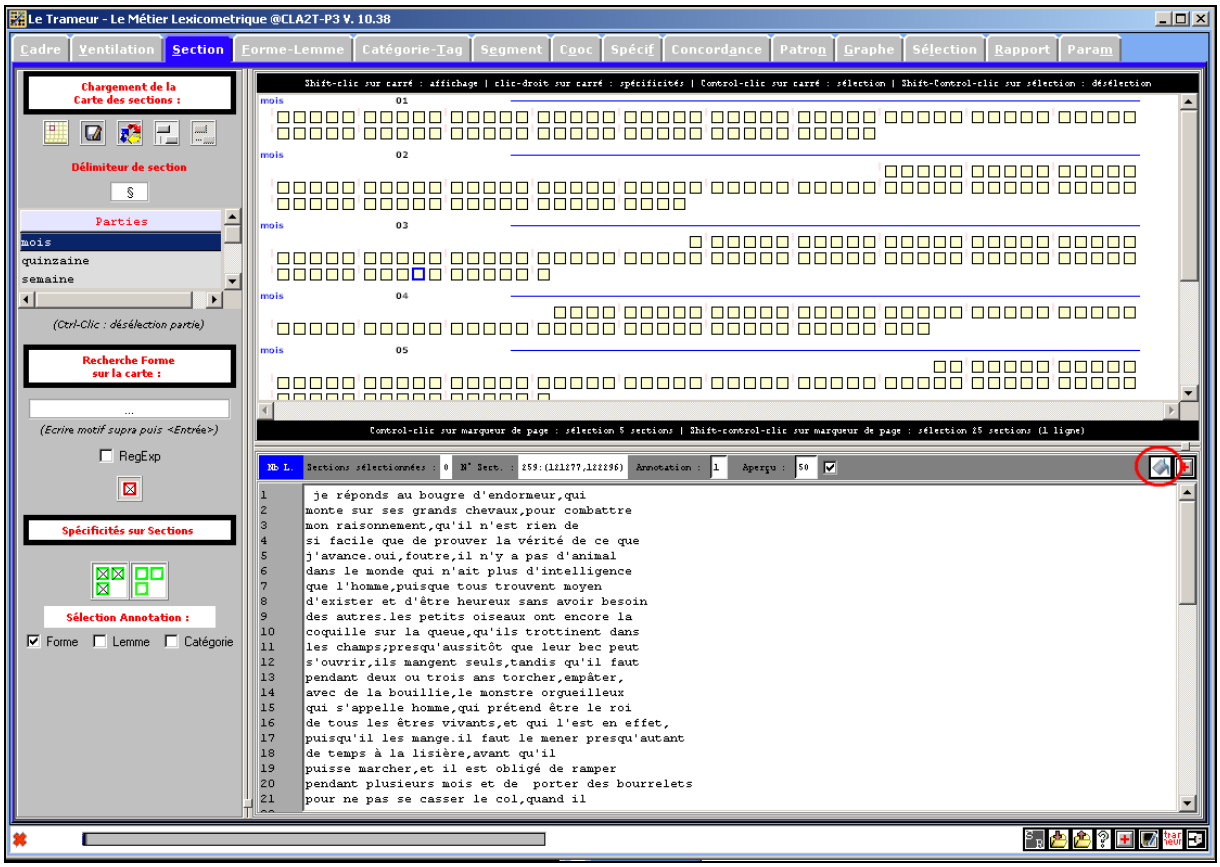

Le Trameur - Le Métier Lexicometrique @CLA2T-P3 V. 10.38
Cadre
Ventilation
Section
Forme-Lemme
Catégorie-Tag
Segment
Cooc
Spécif
Concordance
Patron
Graphe
Sélection
Rapport
Param
Chargement de la Carte des sections :
Délimiteur de section
Parties
mois
quinzaine
semaine
(Ctrl-Clic : désélection partie)
Recherche Forme sur la carte :
(Ecrire motif supra puis <Entrée>)
RegExp
Spécificités sur Sections
Sélection Annotation :
Forme
Lemme
Catégorie
Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : désélection
Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)
je réponds au bougre d'endormeur,qui
monte sur ses grands chevaux,pour combattre
mon raisonnement,qu'il n'est rien de
si facile que de prouver la vérité de ce que
j'avance.oui,foutre,il n'y a pas d'animal
dans le monde qui n'ait plus d'intelligence
que l'homme,puisque tous trouvent moyen
d'exister et d'être heureux sans avoir besoin
des autres.les petits oiseaux ont encore la
coquille sur la queue,qu'ils trottinent dans
les champs;presqu'aussitôt que leur bec peut
s'ouvrir,ils mangent seuls,tandis qu'il faut
pendant deux ou trois ans torcher,empâter,
avec de la bouillie,le monstre orgueilleux
qui s'appelle homme,qui prétend être le roi
de tous les êtres vivants,et qui l'est en effet,
puisqu'il les mange.il faut le mener presqu'autant
de temps à la lisière,avant qu'il
puisse marcher,et il est obligé de ramper
pendant plusieurs mois et de porter des bourrelets
pour ne pas se casser le col,quand il

Figure 50 : *Le Trameur* Editeur SECTION

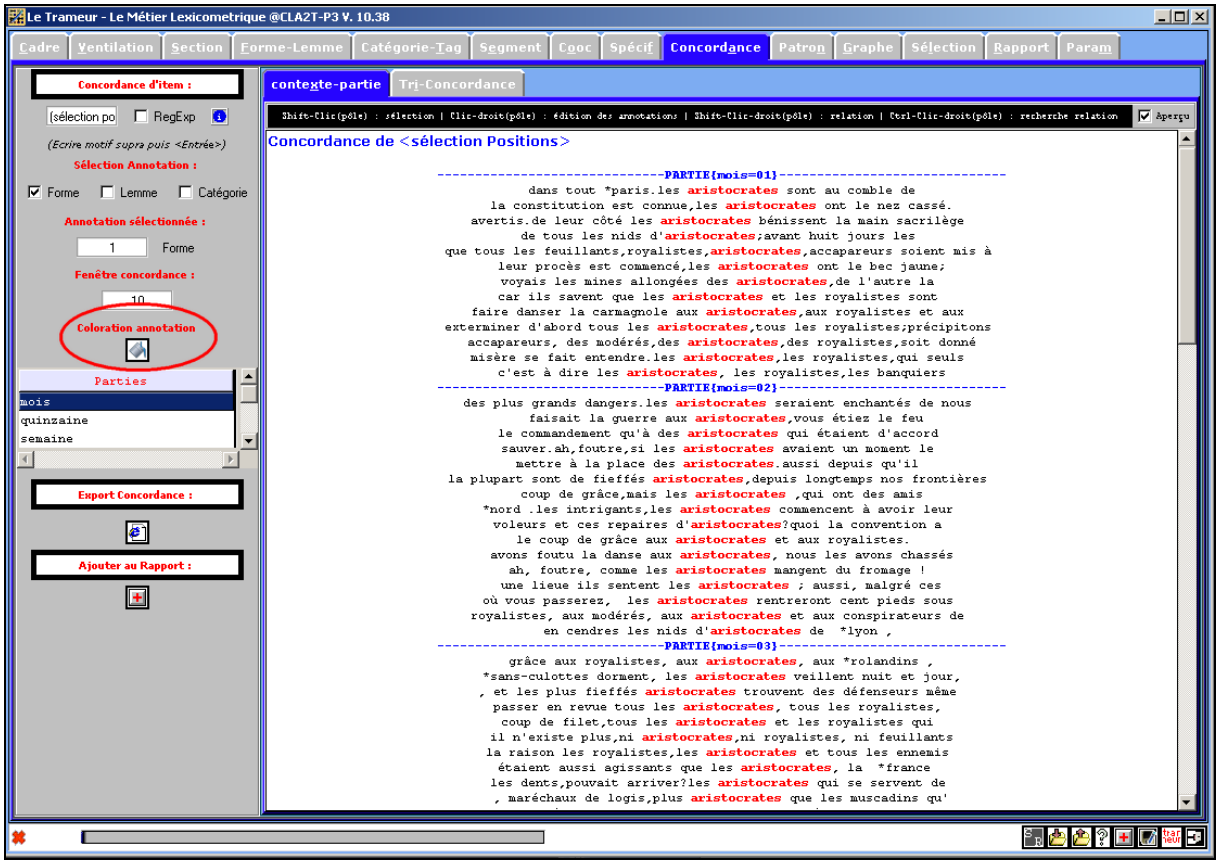

Le Trameur - Le Métier Lexicometrique @CLA2T-P3 V. 10.38
Cadre
Ventilation
Section
Forme-Lemme
Catégorie-Tag
Segment
Cooc
Spécif
Concordance
Patron
Graphe
Sélection
Rapport
Param
Concordance d'item :
RegExp
(Ecrire motif supra puis <Entrée>)
Sélection Annotation :
Forme
Lemme
Catégorie
Annotation sélectionnée :
Forme
Fenêtre concordance :
Coloration annotation
Parties
mois
quinzaine
semaine
Export Concordance :
Ajouter au Rapport :
contexte-partie
Tri-Concordance
Aperçu
Concordance de <sélection Positions>
----------PARTIE{mois=01}----------
dans tout *paris.les aristocrates sont au comble de
la constitution est connue,les aristocrates ont le nez cassé.
avertis.de leur côté les aristocrates bénissent la main sacrilège
de tous les nids d'aristocrates;avant huit jours les
que tous les feuillants,royalistes,aristocrates,accapareurs soient mis à
leur procès est commencé,les aristocrates ont le bec jaune;
voyais les mines allongées des aristocrates,de l'autre la
car ils savent que les aristocrates et les royalistes sont
faire danser la carmagnole aux aristocrates,aux royalistes et aux
exterminer d'abord tous les aristocrates,tous les royalistes;précipitons
accapareurs, des modérés,des aristocrates,des royalistes,soit donné
misère se fait entendre.les aristocrates,les royalistes,qui seuls
c'est à dire les aristocrates, les royalistes,les banquiers
----------PARTIE{mois=02}----------
des plus grands dangers.les aristocrates seraient enchantés de nous
faisait la guerre aux aristocrates,vous étiez le feu
le commandement qu'à des aristocrates qui étaient d'accord
sauver.ah,foutre,si les aristocrates avaient un moment le
mettre à la place des aristocrates.aussi depuis qu'il
la plupart sont de fieffés aristocrates,depuis longtemps nos frontières
coup de grâce,mais les aristocrates ,qui ont des amis
*nord .les intrigants,les aristocrates commencent à avoir leur
voleurs et ces repaires d'aristocrates?quoi la convention a
le coup de grâce aux aristocrates et aux royalistes.
avons foutu la danse aux aristocrates, nous les avons chassés
ah, foutre, comme les aristocrates mangent du fromage !
une lieue ils sentent les aristocrates ; aussi, malgré ces
où vous passerez, les aristocrates rentreront cent pieds sous
royalistes, aux modérés, aux aristocrates et aux conspirateurs de
en cendres les nids d'aristocrates de *lyon ,
----------PARTIE{mois=03}----------
grâce aux royalistes, aux aristocrates, aux *rolandins ,
*sans-culottes dorment, les aristocrates veillent nuit et jour,
, et les plus fieffés aristocrates trouvent des défenseurs même
passer en revue tous les aristocrates, tous les royalistes,
coup de filet,tous les aristocrates et les royalistes qui
il n'existe plus,ni aristocrates,ni royalistes, ni feuillants
la raison les royalistes,les aristocrates et tous les ennemis
étaient aussi agissants que les aristocrates, la *france
les dents,pouvait arriver?les aristocrates qui se servent de
, maréchaux de logis,plus aristocrates que les muscadins qu'

Figure 51 : *Le Trameur* Editeur CONCORDANCE

L'activation du bouton déclenche le menu suivant :

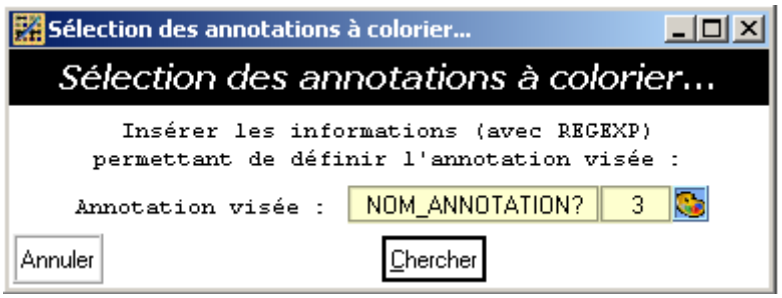


Figure 52 : *Le Trameur* Sélection des annotations à colorier (1)

Ce menu permet de choisir la valeur de l'annotation à colorier, le numéro de l'annotation correspondante et la couleur à utiliser : tous les niveaux d'annotation sont accessibles pour cette fonctionnalité.

Par exemple, dans la figure suivante, on choisit l'étiquette « NOM » associée à l'annotation n°3 et la couleur visée par la palette de choix :

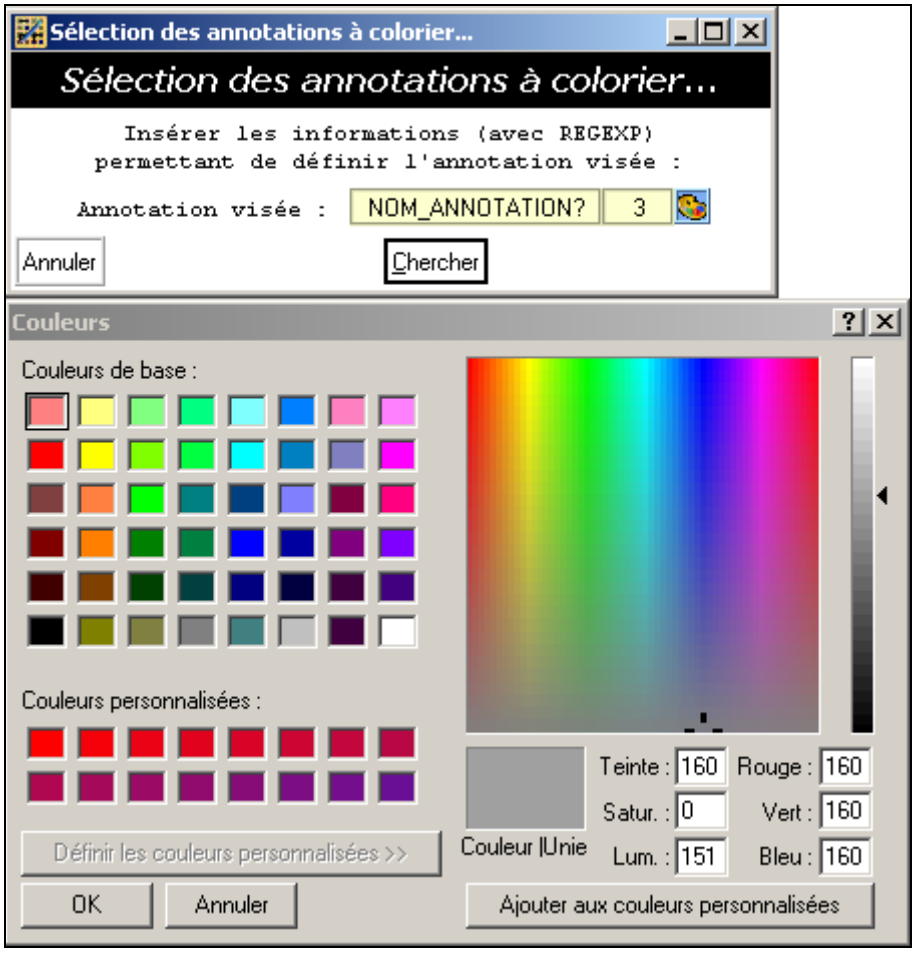


Figure 53 : *Le Trameur* Sélection des annotations à colorier (2)

Le résultat dans les différents éditeurs est le suivant :

- **Editeur de la Carte des sections**

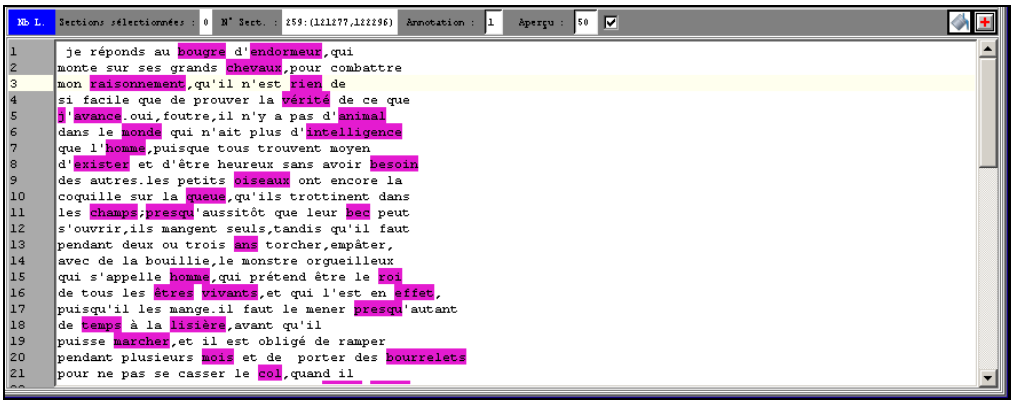

Nb L.
Sections sélectionnées : 0
N° Sect. : 259:(121277,122296)
Annotation : 1
Aperçu : 50
je réponds au bougre d'endormeur,qui
monte sur ses grands chevaux,pour combattre
mon raisonnement,qu'il n'est rien de
si facile que de prouver la vérité de ce que
l'avance.oui,foutre,il n'y a pas d'animal
dans le monde qui n'ait plus d'intelligence
que l'homme,puisque tous trouvent moyen
d'exister et d'être heureux sans avoir besoin
des autres.les petits oiseaux ont encore la
coquille sur la queue,qu'ils trottinent dans
les champs;presqu'aussitôt que leur bec peut
s'ouvrir,ils mangent seuls,tandis qu'il faut
pendant deux ou trois ans torcher,empâter,
avec de la bouillie,le monstre orgueilleux
qui s'appelle homme,qui prétend être le roi
de tous les êtres vivants,et qui l'est en effet,
puisqu'il les mange.il faut le mener presqu'autant
de temps à la lisière,avant qu'il
puisse marcher,et il est obligé de ramper
pendant plusieurs mois et de porter des bourrelets
pour ne pas se casser le col,quand il

Figure 54 : *Le Trameur* Marquage des annotations sur les items édités (1)

- **Editeur du *Cadre***

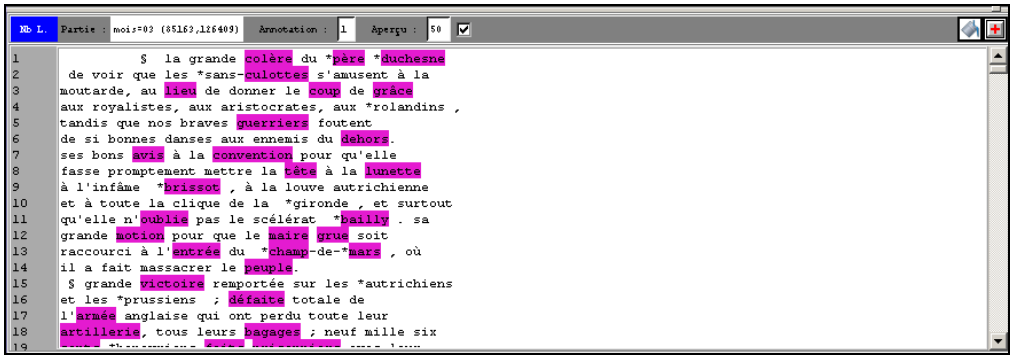

Nb L.
Partie : mois=03 (85163,128409)
Annotation : 1
Aperçu : 50
$ la grande colère du *père *duchesne
de voir que les *sans-culottes s'amusent à la
moutarde, au lieu de donner le coup de grâce
aux royalistes, aux aristocrates, aux *rolandins ,
tandis que nos braves guerriers foutent
de si bonnes danses aux ennemis du dehors.
ses bons avis à la convention pour qu'elle
fasse promptement mettre la tête à la lunette
à l'infâme *brissot , à la louve autrichienne
et à toute la clique de la *gironde , et surtout
qu'elle n'oublie pas le scélérat *bailly . sa
grande motion pour que le maire grue soit
raccourci à l'entrée du *champ-de-*mars , où
il a fait massacrer le peuple.
$ grande victoire remportée sur les *autrichiens
et les *prussiens ; défaite totale de
l'armée anglaise qui ont perdu toute leur
artillerie, tous leurs bagages ; neuf mille six

Figure 55 : *Le Trameur* Marquage des annotations sur les items édités (2)

- **Editeur du concordancier**

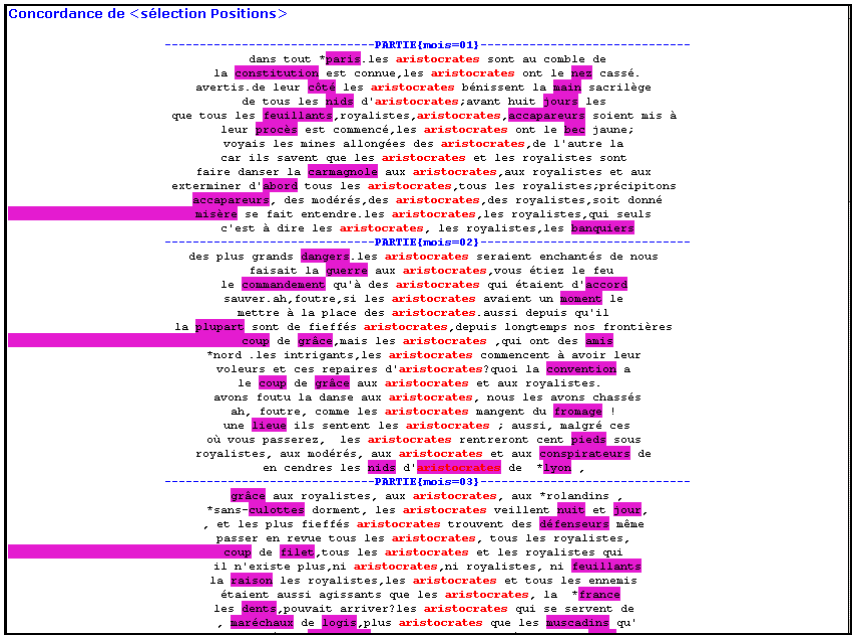

Concordance de <sélection Positions>
----------PARTIE{mois=01}----------
dans tout *paris.les aristocrates sont au comble de
la constitution est connue,les aristocrates ont le nez cassé.
avertis.de leur côté les aristocrates bénissent la main sacrilège
de tous les nids d'aristocrates;avant huit jours les
que tous les feuillants,royalistes,aristocrates,accapareurs soient mis à
leur procès est commencé,les aristocrates ont le bec jaune;
voyais les mines allongées des aristocrates,de l'autre la
car ils savent que les aristocrates et les royalistes sont
faire danser la carmagnole aux aristocrates,aux royalistes et aux
exterminer d'abord tous les aristocrates,tous les royalistes;précipitons
accapareurs, des modérés,des aristocrates,des royalistes,soit donné
misère se fait entendre.les aristocrates,les royalistes,qui seuls
c'est à dire les aristocrates, les royalistes,les banquiers
----------PARTIE{mois=02}----------
des plus grands dangers.les aristocrates seraient enchantés de nous
faisait la guerre aux aristocrates,vous étiez le feu
le commandement qu'à des aristocrates qui étaient d'accord
sauver.ah,foutre,si les aristocrates avaient un moment le
mettre à la place des aristocrates.aussi depuis qu'il
la plupart sont de fieffés aristocrates,depuis longtemps nos frontières
coup de grâce,mais les aristocrates ,qui ont des amis
*nord .les intrigants,les aristocrates commencent à avoir leur
voleurs et ces repaires d'aristocrates?quoi la convention a
le coup de grâce aux aristocrates et aux royalistes.
avons foutu la danse aux aristocrates, nous les avons chassés
ah, foutre, comme les aristocrates mangent du fromage !
une lieue ils sentent les aristocrates ; aussi, malgré ces
où vous passerez, les aristocrates rentreront cent pieds sous
royalistes, aux modérés, aux aristocrates et aux conspirateurs de
en cendres les nids d'aristocrates de *lyon ,
----------PARTIE{mois=03}----------
grâce aux royalistes, aux aristocrates, aux *rolandins ,
*sans-culottes dorment, les aristocrates veillent nuit et jour,
, et les plus fieffés aristocrates trouvent des défenseurs même
passer en revue tous les aristocrates, tous les royalistes,
coup de filet,tous les aristocrates et les royalistes qui
il n'existe plus,ni aristocrates,ni royalistes, ni feuillants
la raison les royalistes,les aristocrates et tous les ennemis
étaient aussi agissants que les aristocrates, la *france
les dents,pouvait arriver?les aristocrates qui se servent de
, maréchaux de logis,plus aristocrates que les muscadins qu'

Figure 56 : *Le Trameur* Marquage des annotations sur les items édités (3)

On peut « empiler » les couleurs et marquer plusieurs annotations par un système de couleurs variées dans la zone d'édition : on choisit ci-dessous une autre valeur d'annotation (ADJ) et on applique la coloration dans l'éditeur de la carte des sections :

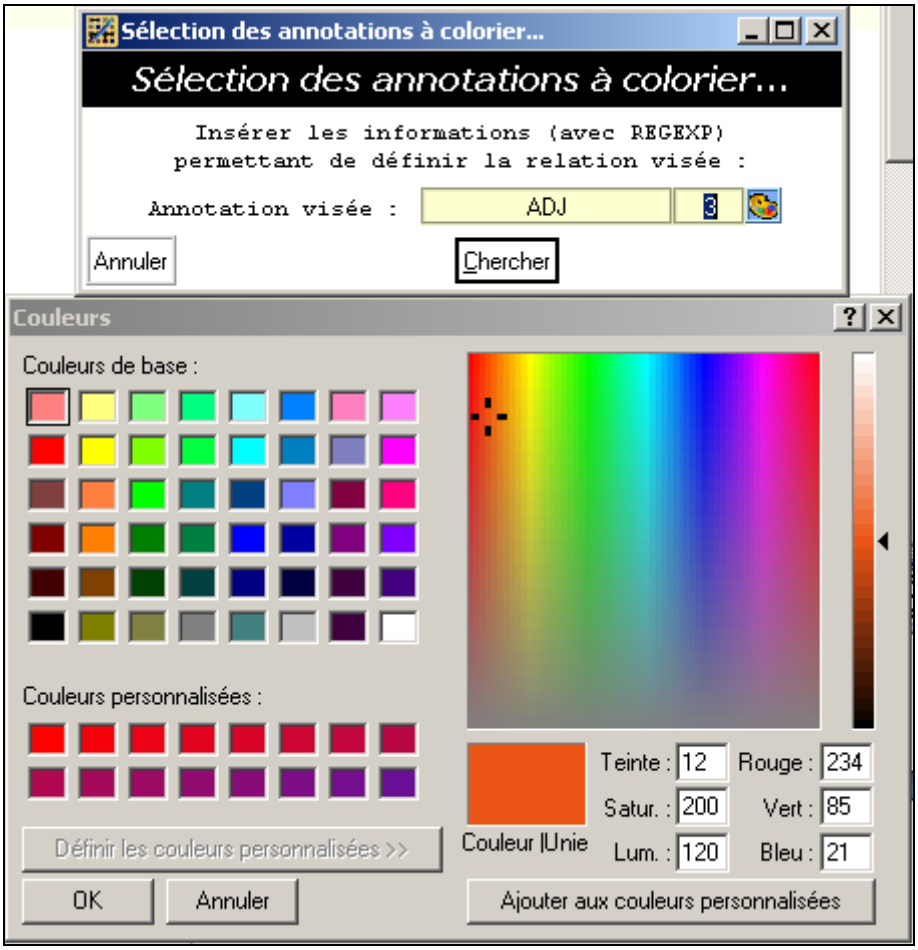


Figure 57 : *Le Trameur* Sélection d'une nouvelle annotation à colorier

Le résultat dans les éditeurs est le suivant :

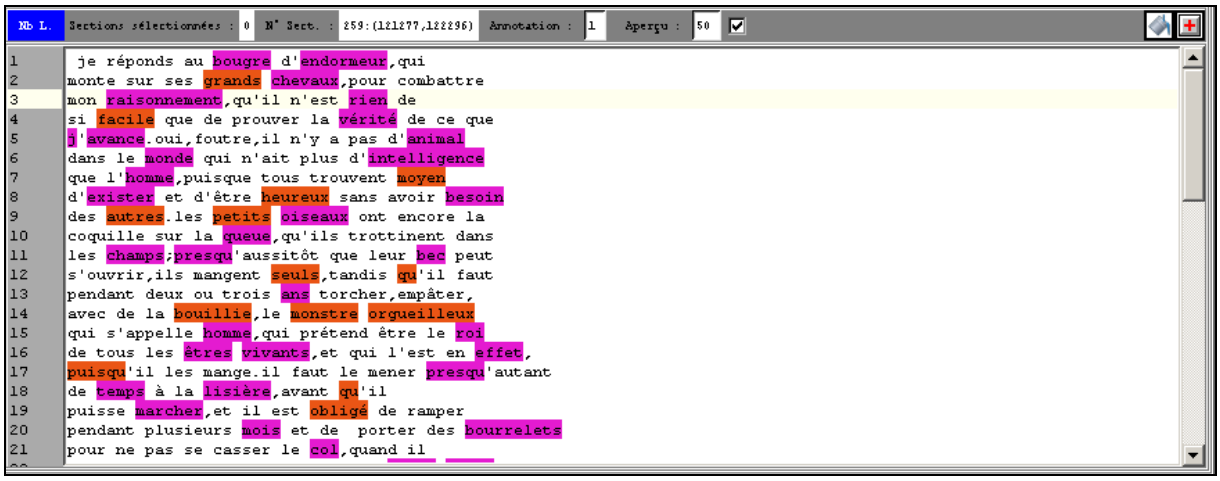

Nb L.
Sections sélectionnées : 0
N° Sect. : 259:(121277,122296)
Annotation : 1
Aperçu : 50
je réponds au bougre d'endormeur,qui
monte sur ses grands chevaux,pour combattre
mon raisonnement,qu'il n'est rien de
si facile que de prouver la vérité de ce que
j'avance.oui,foutre,il n'y a pas d'animal
dans le monde qui n'ait plus d'intelligence
que l'homme,puisque tous trouvent moyen
d'exister et d'être heureux sans avoir besoin
des autres.les petits oiseaux ont encore la
coquille sur la queue,qu'ils trottinent dans
les champs;presqu'aussitôt que leur bec peut
s'ouvrir,ils mangent seuls,tandis qu'il faut
pendant deux ou trois ans torcher,empâter,
avec de la bouillie,le monstre orgueilleux
qui s'appelle homme,qui prétend être le roi
de tous les êtres vivants,et qui l'est en effet,
puisqu'il les mange.il faut le mener presqu'autant
de temps à la lisière,avant qu'il
puisse marcher,et il est obligé de ramper
pendant plusieurs mois et de porter des bourrelets
pour ne pas se casser le col,quand il

Figure 58 : *Le Trameur* Marquage des annotations sur les items édités

14.2 Annotation de type « relation »

Si l'annotation recherchée est de type « relation » entre 2 items (*cf* Graphe de relations *infra*), la source et la cible de la relation sont marquées dans l'éditeur.

Pour illustrer cette fonctionnalité, on utilise dans l'exemple qui suit la base textométrique intitulée **Base "Corpus de contes français"** (Fairy Tales Corpus, FTC) construite à partir des ressources développées par Ismaïl El Maarouf (base décrite en détail dans la partie « Graphes de relations *infra*).

On considère par exemple ici la section suivante de cette base :

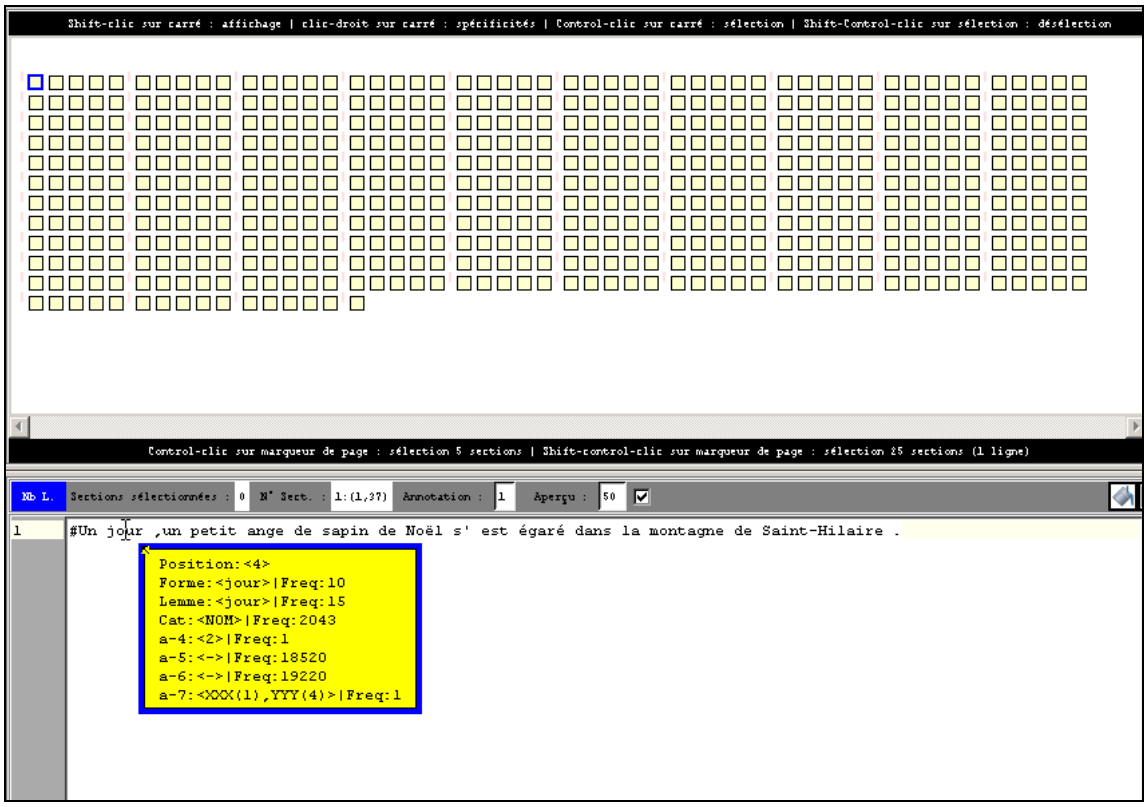

Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : désélection
Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)
Nb L.
Sections sélectionnées :
N° Sect. :
Annotation :
Aperçu :
#Un jour ,un petit ange de sapin de Noël s' est égaré dans la montagne de Saint-Hilaire .
Position:<4>
Forme:<jour>|Freq:10
Lemme:<jour>|Freq:15
Cat:<NOM>|Freq:2043
a-4:<2>|Freq:1
a-5:<->|Freq:18520
a-6:<->|Freq:19220
a-7:<XXX(1),YYY(4)>|Freq:1

Figure 59 : *Le Trameur* Edition d'une annotation de type « relation »

Dans celle-ci, l'item en position n°4 (la forme « jour ») possède des annotations suivantes :

- **Annotation n°1 (forme)** : `jour`
- **Annotation n°2 (lemme)** : `jour`
- **Annotation n°3 (POS)** : `NOM`
- **Annotation n°4 (identifiant)** : `2` (cette annotation permet d'identifier chaque item par un identifiant unique. Remarque : la position de chaque item est une autre manière d'identifier les items de la ***Trame***)
- **Annotation n°5** : annotation référentielle non définie
- **Annotation n°6** : annotation sémantique non définie
- **Annotation n°7** : `XXX(1), YYY(4)`

Cette dernière annotation induit des relations sur l'item courant : par exemple celle notée YYY(4). Cette valeur d'annotation signifie que l'item courant est en relation de type YYY avec un autre item de la ***Trame*** (celui dont l'identifiant est 4)

Si on souhaite afficher par exemple les relations de type YYY, l'activation du bouton déclenche le menu suivant :

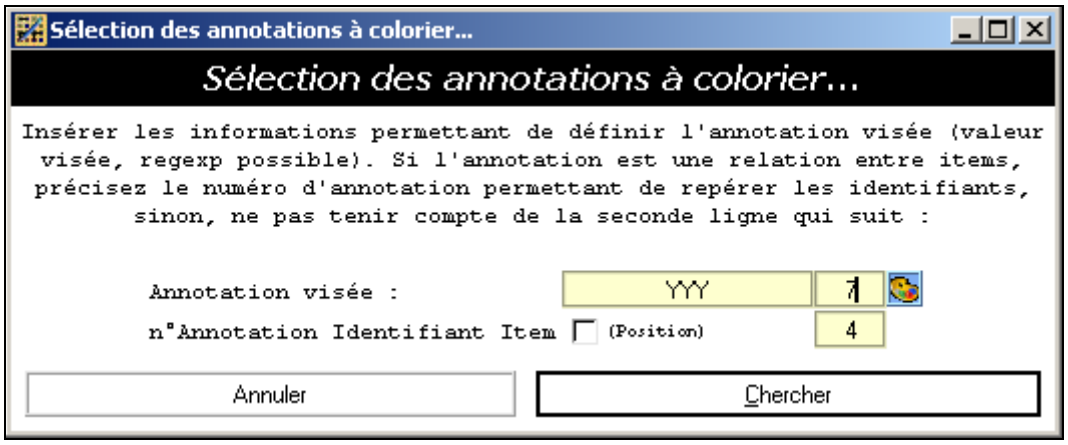

Sélection des annotations à colorier...
Sélection des annotations à colorier...
Insérer les informations permettant de définir l'annotation visée (valeur visée, regexp possible). Si l'annotation est une relation entre items, précisez le numéro d'annotation permettant de repérer les identifiants, sinon, ne pas tenir compte de la seconde ligne qui suit :
Annotation visée :
YYY
7
n°Annotation Identifiant Item
(Position)
4
Annuler
Chercher

Figure 60 : *Le Trameur* Sélection d'une relation à colorier

La première ligne permet comme précédemment de définir la valeur de l'annotation visée (YYY stockée dans l'annotation n°7)

La seconde permet de sélectionner l'annotation dans laquelle sont enregistrés les identifiants des items :

- soit l'identifiant est stocké dans un niveau d'annotation particulier : c'est le cas ici, l'annotation n°4 a été construite pour enregistrer un identifiant unique sur chaque item de type non délimiteur
- soit il correspond à la position de l'item (cette identification par la position peut être considérée comme l'identification par défaut et par construction de chacun des items de la ***Trame***)

Le résultat produit est cette fois-ci le suivant :

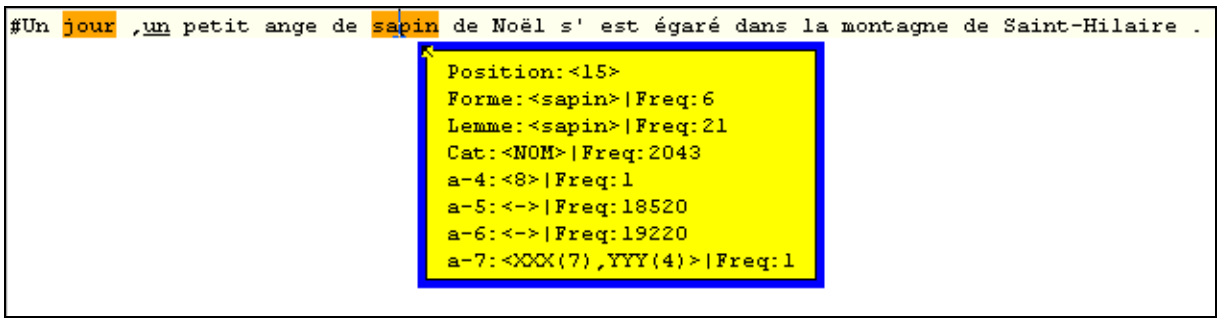


Figure 61 : *Le Trameur* Marquage des annotations de type « relation » sur les items édités

L'item en position n°4 est colorié (source de la relation YYY(4))
L'item associé à identifiant n°4 (cible de la relation) est souligné (la forme « un »)
Dans cet exemple, l'item en position n°15 a une relation similaire vers la même cible.

15 Vocabulaire spécifique

L'analyse des spécificités permet de porter un jugement sur la fréquence de chacune des unités textuelles dans chacune des parties du corpus[13].

Tous les résultats d'un calcul de spécificités sont insérés dans l'onglet **SPECIF**. Ils sont associés à des processus de calcul disponibles dans différents onglets.

Tous les modules intégrant un calcul de spécificité peuvent déclencher ce calcul sur tous les niveaux d'annotation disponibles dans la base en cours d'analyse : tous les niveaux d'annotation ajoutés, importés etc. sont accessibles pour le calcul des spécificités.

Localisation du paramétrage de l'annotation pour le calcul des spécificités et modules associés :

Onglet SECTION
- calcul des spécificités des sections sélectionnées
- calcul des spécificités des sections contenant un motif donné (i.e cooccurrent du motif visé)

Onglet SPECIF
- calcul des spécificités sur une partie donnée (sélectionnée via PCLC)
- spécificités totales sur une partition (TFG)
- AFC

Onglet SELECTION - sélection sur ***Cadre*** (i.e sélection de parties)
- calcul des spécifs sur une partie donnée

Le paramétrage de l'annotation visée est similaire dans ses trois onglets :

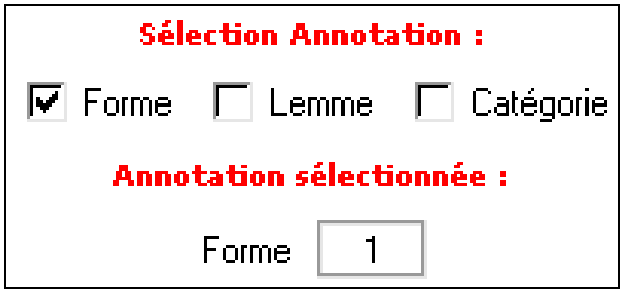


Figure 62 : *Le Trameur* Sélection d'annotation pour calcul de spécificités

3 cases à cocher permettent de sélectionner les annotations : forme (annotation n°1), lemme (annotation n°2) ou catégorie (annotation n°3). Pour les autres niveaux d'annotation, il suffit de donner son numéro dans la zone de saisie (puis entrée), le nom de l'annotation choisie sera visible en regard de son numéro : dans la figure précédente, l'annotation n°1 (forme) est sélectionnée.

[13] Sur la méthode des spécificités on consultera par exemple : (Lafon, 1984) ou en ligne : http://weblex.ens-lsh.fr/doc/weblex/contraste.html#specif

Toutes les opérations (concordance, ventilation, etc) activables sur les résultats de spécificités sont disponibles sur tous les niveaux d'annotation (*cf infra*).

Les résultats d'un calcul de spécificités sont insérés dans 2 sous-onglets de l'onglet **SPECIF**, chacun d'eux permet d'activer des opérations similaires :

a. sous-onglet **SPECIF**:

Les résultats sont affichés ici en format texte. Les actions disponibles sur les items affichés sont activables par des raccourcis clavier (encadrés en rouge) :

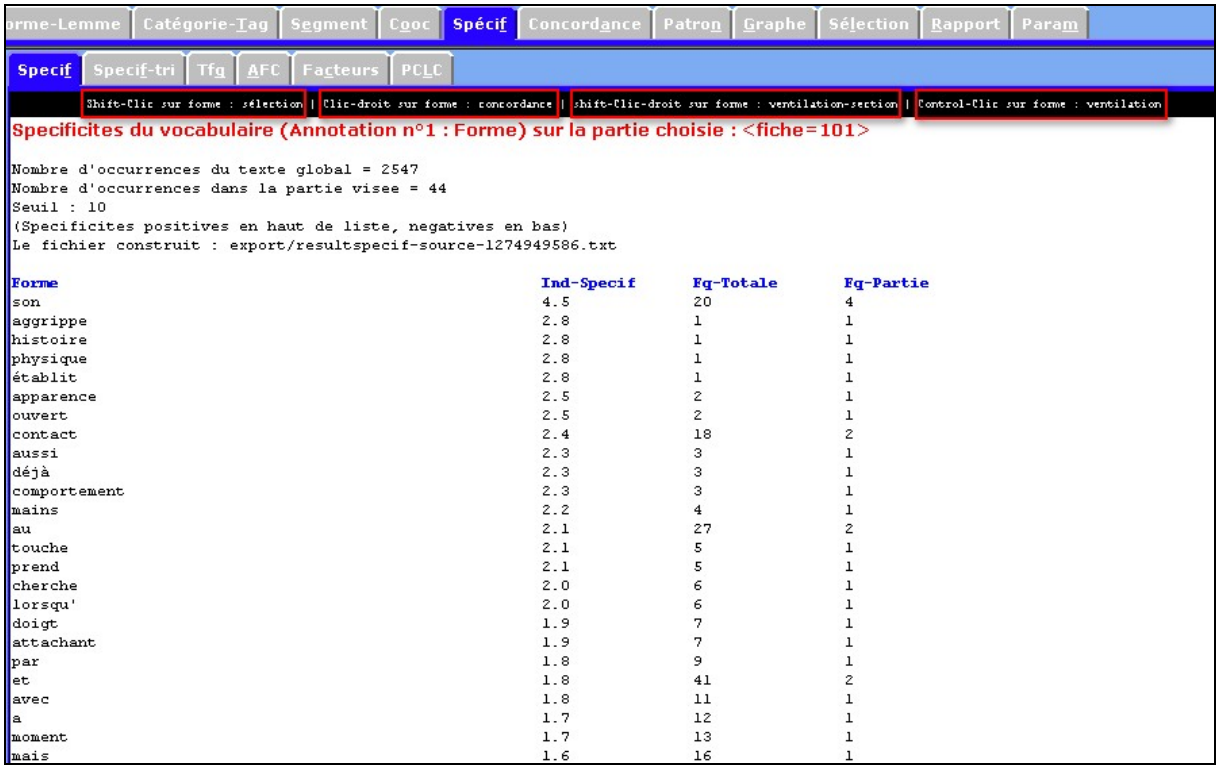
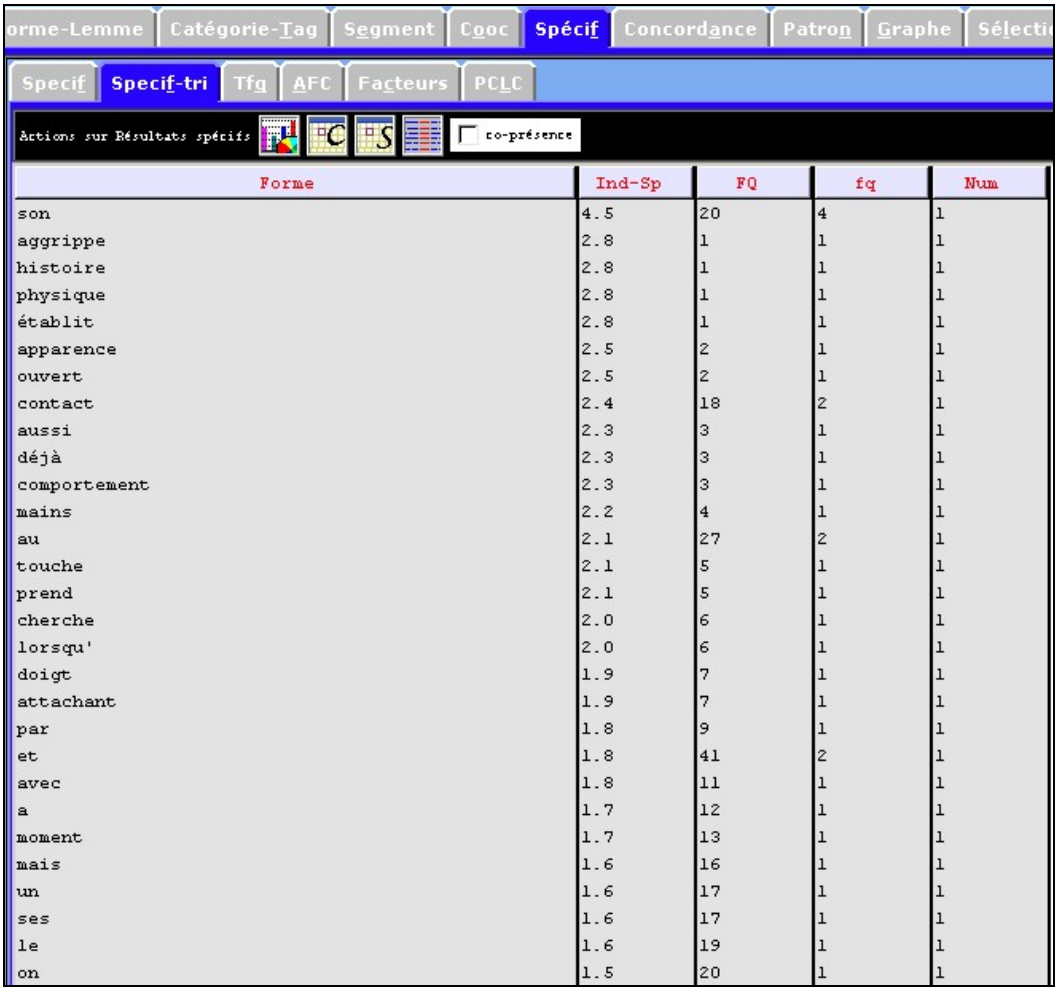


Figure 63 : *Le Trameur* Opérations disponibles sur les items issus d'un calcul de spécificités

- *Sélection* : tous les items ayant la forme visée sont sélectionnés (*cf Gestionnaire de Sélection*)
- *Concordance* : calcul d'une concordance sur toutes les positions des items ayant la forme visée
- *Ventilation-section* : ventilation sur la carte des sections de toutes les positions des items ayant la forme visée
- *Ventilation*: ventilation sur le **CADRE** de toutes les positions des items ayant la forme visée

b. sous-onglet **SPECIF-TRI**

Dans cet onglet, les résultats affichés sous forme de colonnes donnent à voir pour chaque item : l'item visé, son indice de spécificité, sa fréquence totale, sa fréquence dans la partie et le numéro d'annotation de l'item (1 : forme, 2 : lemme, 3 : catégorie, etc.).



Forme	Ind-Sp	FQ	fq	Num
son	4.5	20	4	1
aggrippe	2.8	1	1	1
histoire	2.8	1	1	1
physique	2.8	1	1	1
établit	2.8	1	1	1
apparence	2.5	2	1	1
ouvert	2.5	2	1	1
contact	2.4	18	2	1
aussi	2.3	3	1	1
déjà	2.3	3	1	1
comportement	2.3	3	1	1
mains	2.2	4	1	1
au	2.1	27	2	1
touche	2.1	5	1	1
prend	2.1	5	1	1
cherche	2.0	6	1	1
lorsqu'	2.0	6	1	1
doigt	1.9	7	1	1
attachant	1.9	7	1	1
par	1.8	9	1	1
et	1.8	41	2	1
avec	1.8	11	1	1
a	1.7	12	1	1
moment	1.7	13	1	1
mais	1.6	16	1	1
un	1.6	17	1	1
ses	1.6	17	1	1
le	1.6	19	1	1
on	1.5	20	1	1

Figure 64 : *Le Trameur* Opérations disponibles sur les items triables issus d'un calcul de spécificités

Dans cet onglet, les processus de calcul sont activables via les boutons disponibles au dessus des listes produites. Il suffit de sélectionner un ou plusieurs items dans la liste puis d'activer le bouton idoine (*ventilation*, *ventilation sur* ***Cadre***, *ventilation sur carte des sections*, *concordance*).

Dans le cas d'une ventilation sur la carte des sections, si l'option « co-présence » est activée, les sections mises au jour seront celles contenant l'ensemble des formes sélectionnées dans la liste (ET logique).

Enfin, dans cet onglet, il est possible de trier chacune des colonnes en cliquant sur le titre de la colonne visée.

16 Graphiques de ventilation

16.1 Ventilation par partie d'un item

L'onglet **VENTILATION** permet d'afficher des graphiques de ventilation de formes, de lemmes ou de catégories dans une partition sélectionnée :

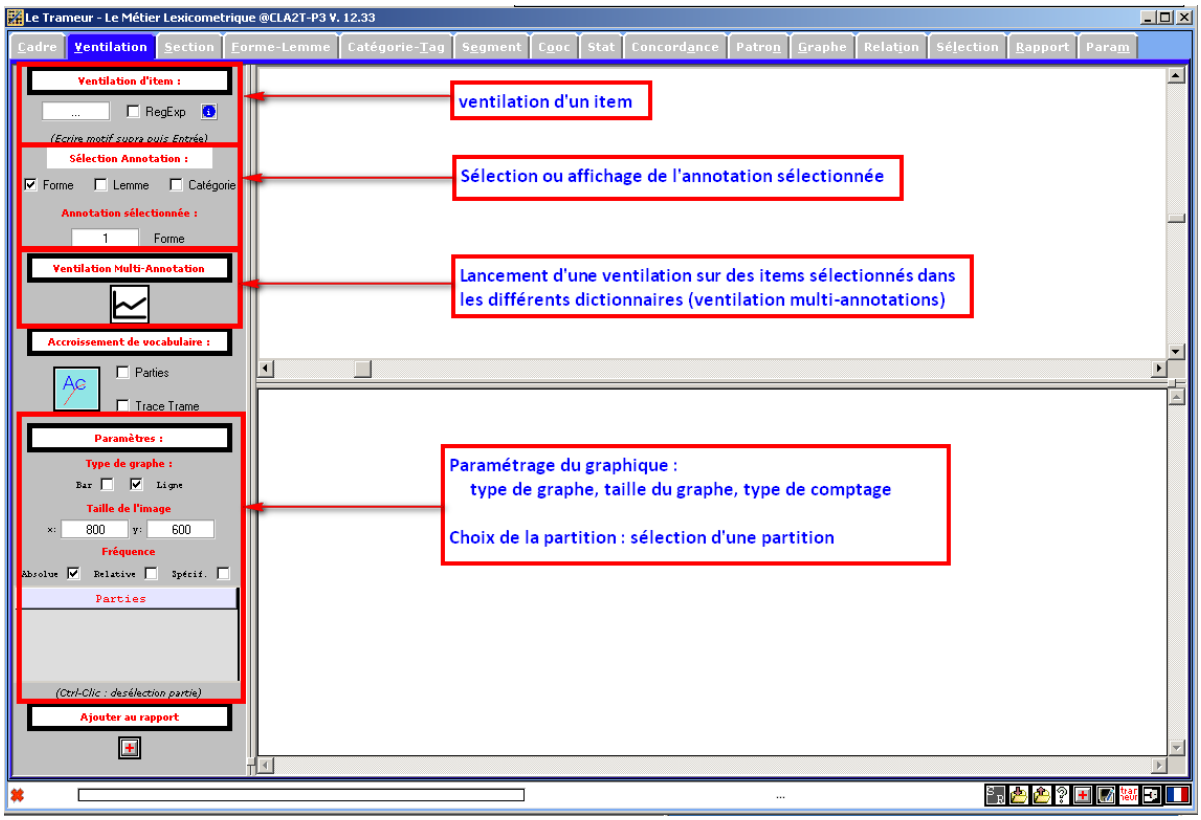


Figure 65 : ***Le Trameur*** **Onglet VENTILATION** paramétrage d'une ventilation

Dans l'exemple suivant, l'item recherché est celui correspondant à la forme « patriote ».

On peut aussi activer la construction du graphe de ventilation en utilisant le mode *RegExp* : la recherche de la forme visée se fait en écrivant un motif exprimé sous la forme d'une expression régulière.

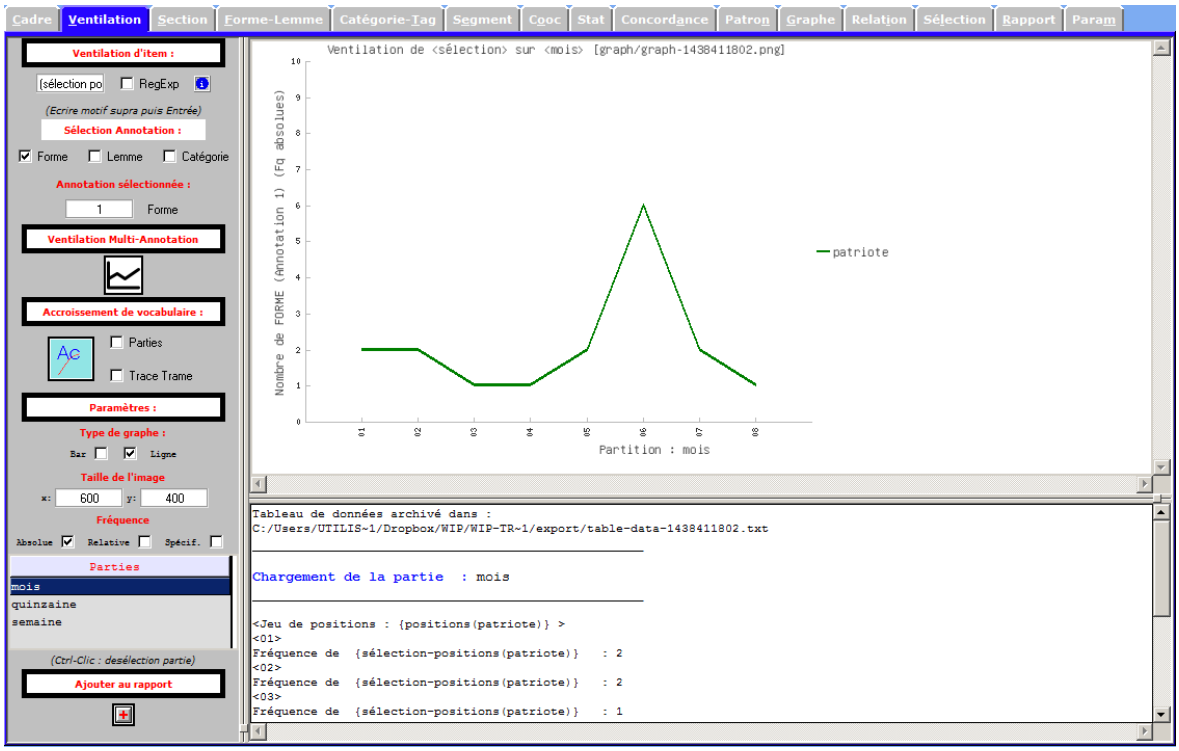


Figure 66 : *Le Trameur* Onglet VENTILATION affichage d'une ventilation (mode ligne)

`\bpatriot.*\b` représente toutes les formes commençant par 'patriot' et qui se termine par n'importe quelle chaîne de caractères (dans notre texte de travail, les chaînes couvertes par cette expressions sont par exemple 'patriote', 'patriotes' etc). Le résultat de la ventilation de ce motif est le suivant :

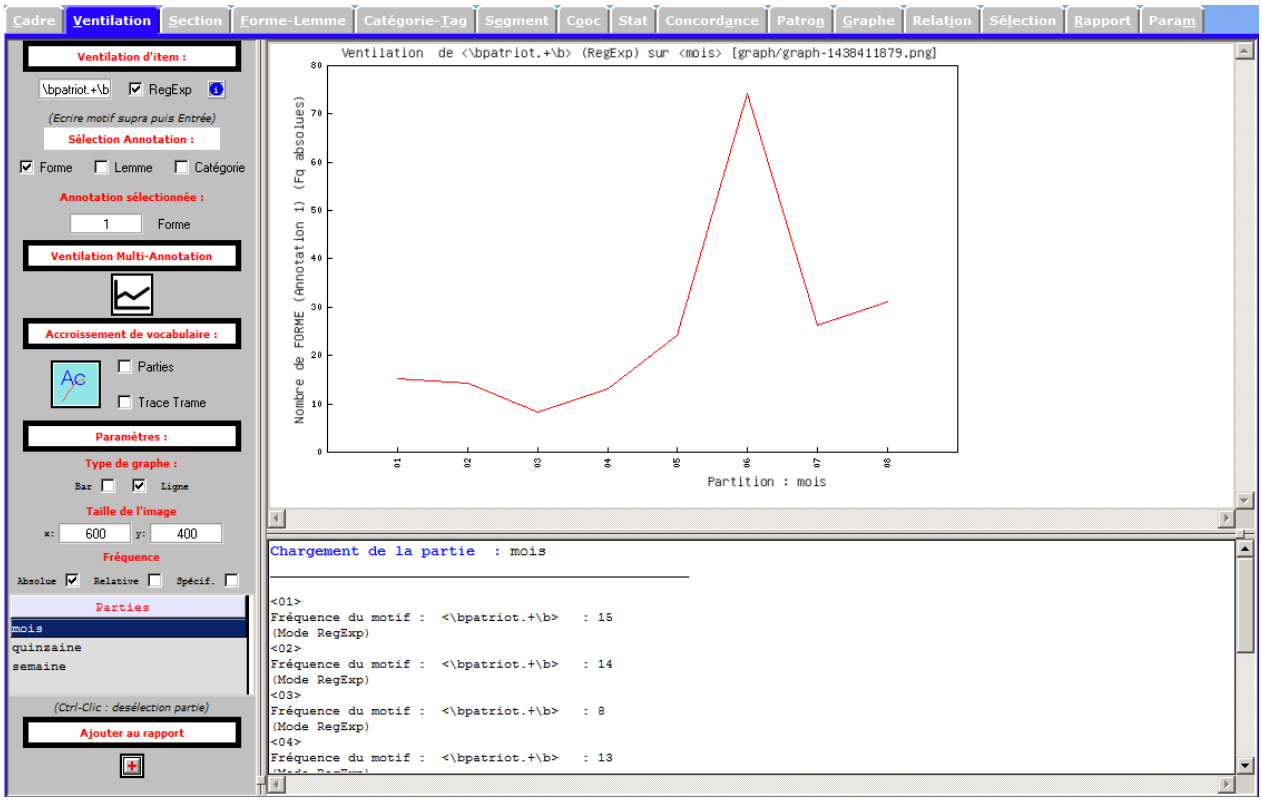


Figure 67 : *Le Trameur* Onglet VENTILATION mode regrexp

16.2 Ventilation par partie d'une sélection d'items

Il est aussi possible de construire des graphiques de ventilation d'une sélection d'item (forme, lemme, catégorie, segment) à partir des listes disponibles. Dans les figures qui suivent, on sélectionne des formes dans le dictionnaire des formes, puis on active la ventilation de ces formes :

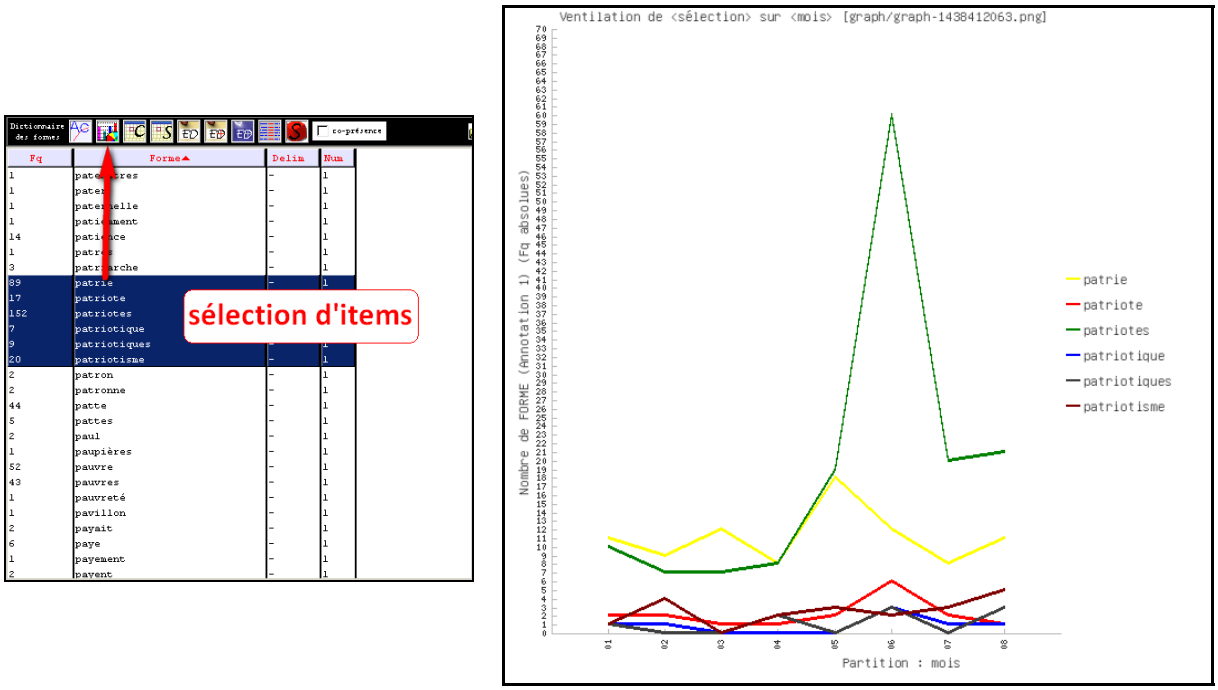


Figure 68 : *Le Trameur* Ventilation d'une sélection d'items

16.3 Courbe d'accroissement du vocabulaire

L'onglet **VENTILATION** permet aussi d'afficher la courbe d'accroissement du vocabulaire sur la base de travail :

- soit sur le corpus complet, en choisissant éventuellement une partition
- soit sur chaque partie de la partition sélectionnée

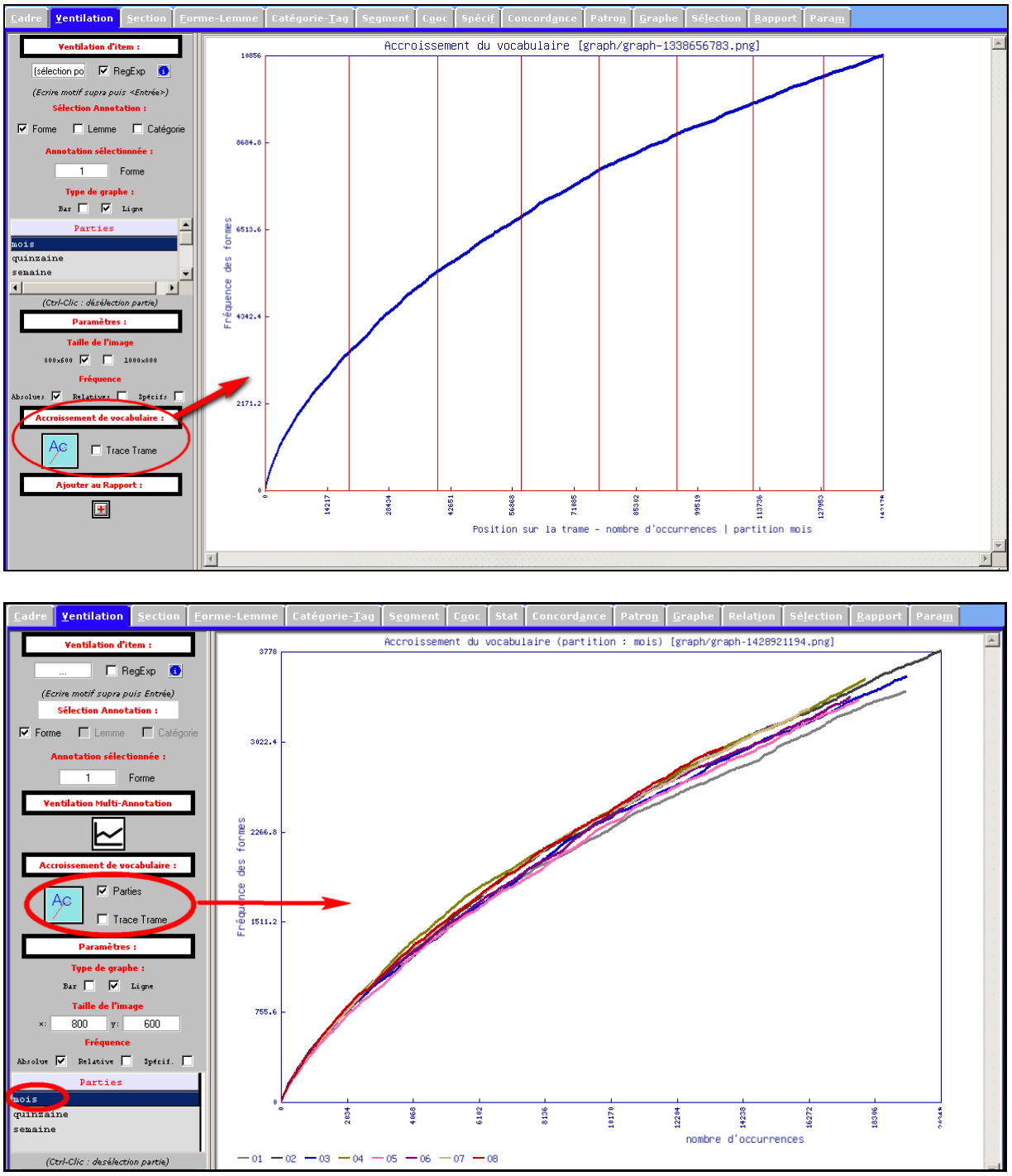


Figure 69 : *Le Trameur* Onglet VENTILATION, courbe d'accroissement du vocabulaire

16.5 Ventilation d'une sélection d'items associés à différents niveaux d'annotation

Il est possible de construire un graphique de ventilation sur une sélection d'items relevant de différents dictionnaires (croisement d'annotation). Après les avoir sélectionnés, l'activation du bonton visible sur la figure suivante déclenche la construction de cette distribution. Ci-dessous, un graphique de ventilation donnant à voir la distribution de la forme « patriotes » et du lemme « patriote » :

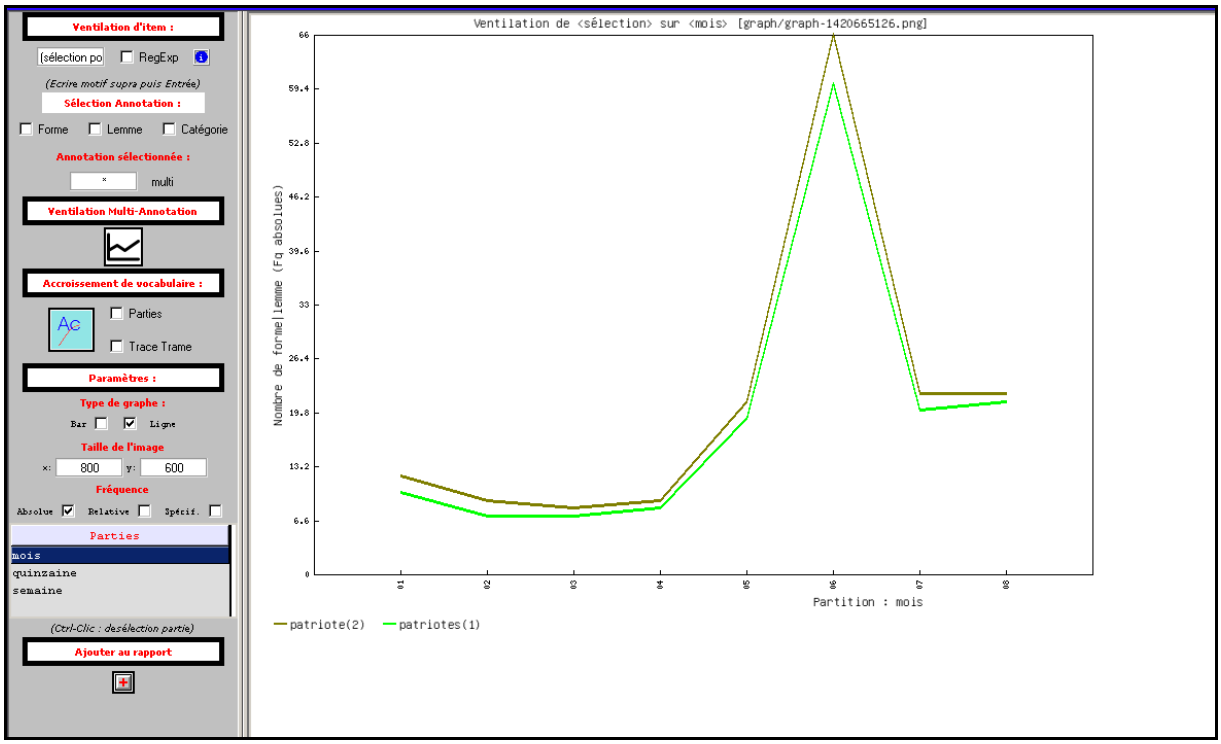


Figure 70 : *Le Trameur* Onglet VENTILATION, ventilation croisée d'annotations

16.4 Ventilation d'une sélection d'items sur une sélection de parties

Le *Gestionnaire de Sélection* permet de sélectionner un certain nombre de parties (*cf infra*). Une fois ces parties sélectionnées, il est possible de construire un graphique de ventilation d'une sélection d'items sur cette sous-partition. Les figures suivantes illustrent le processus en 2 temps :

1. Sélection de parties :

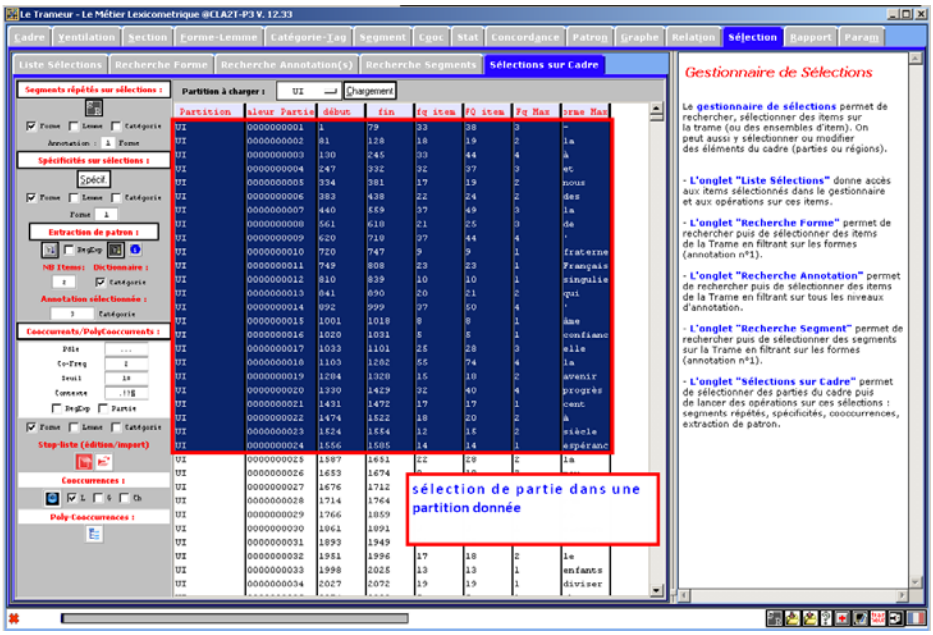


Figure 71 : *Le Trameur* Onglet VENTILATION, ventilation d'une sélection d'items (1)

2. Ventilation d'une sélection d'éléments sur les parties visées :

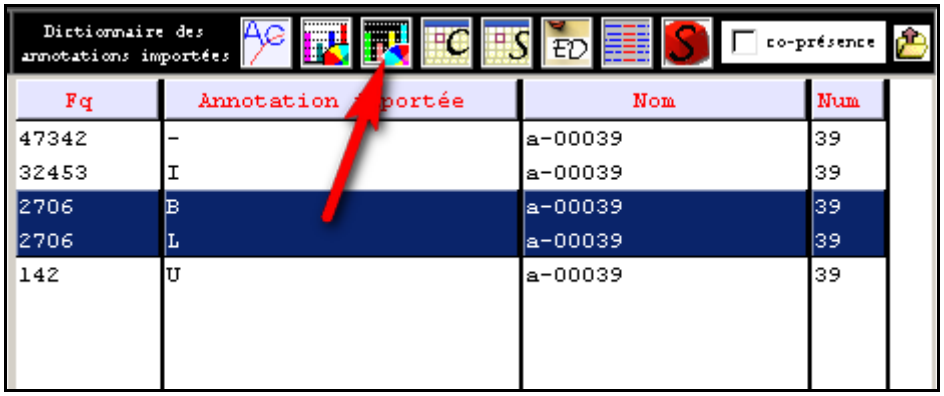


Figure 72 : *Le Trameur* Onglet VENTILATION, ventilation d'une sélection d'items (2)

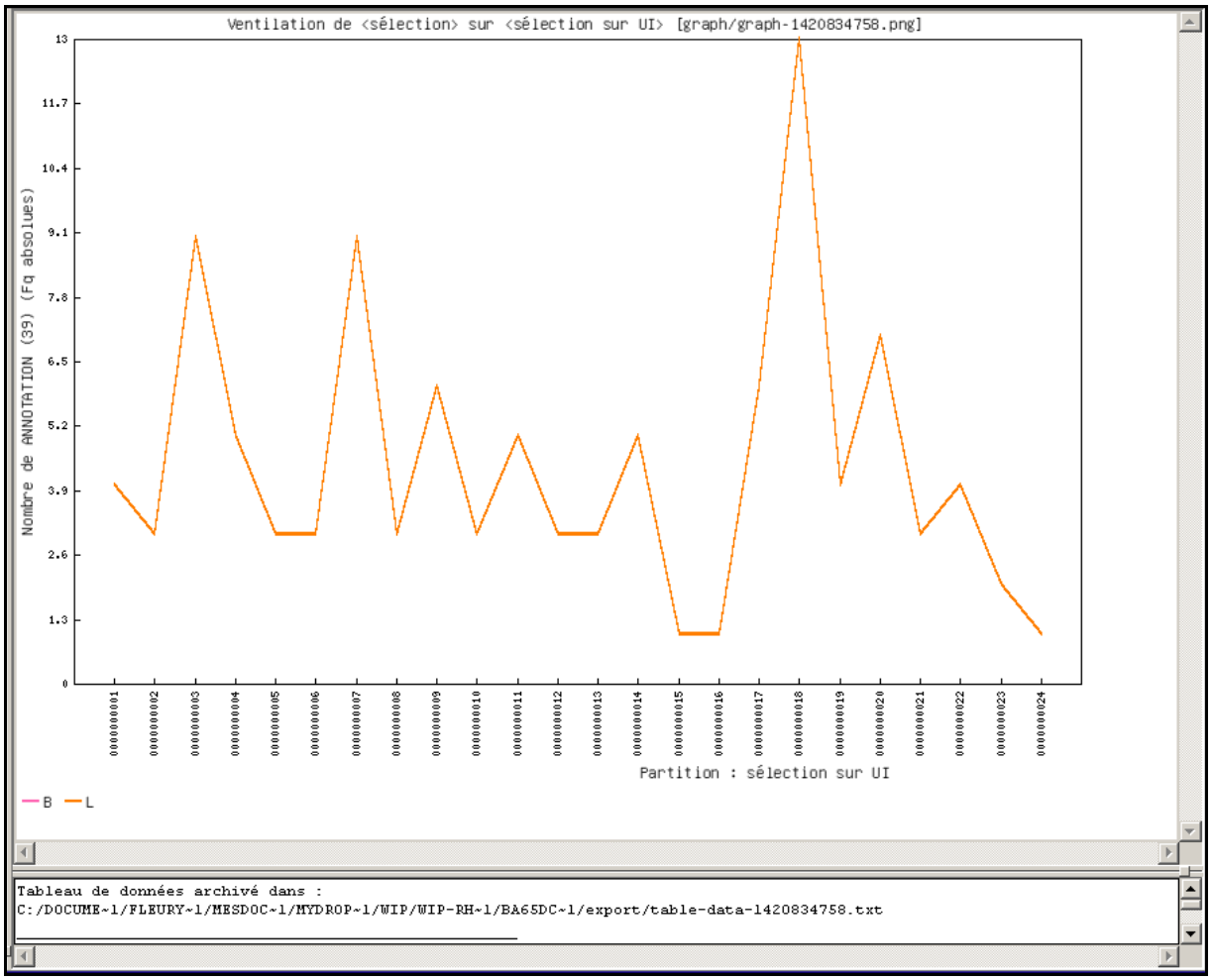


Figure 73 : *Le Trameur* Onglet VENTILATION, ventilation d'une sélection d'items (3)

17 Segments répétés

La fenêtre principale dispose d'un bouton permettant de calculer les segments répétés de la base puis d'importer la liste obtenue dans l'onglet **SEGMENT**. Le calcul des segments est paramétrable (*cf* onglet **PARAM**) : on peut choisir la longueur maximale des segments à rechercher ou leur fréquence minimale. Si ces 2 derniers paramètres ne sont pas renseignés tous les segments de fréquence au moins égale à 2 et de longueur au plus égale à 12 items sont recherchés. Le calcul des segments répétés peut se faire en selectionnant le niveau d'annotation à prendre en compte pour construire ces segments : **Annotation n°1** (forme), **Annotation n°2** (lemme) ou **Annotation n°3** (catégorie) etc.

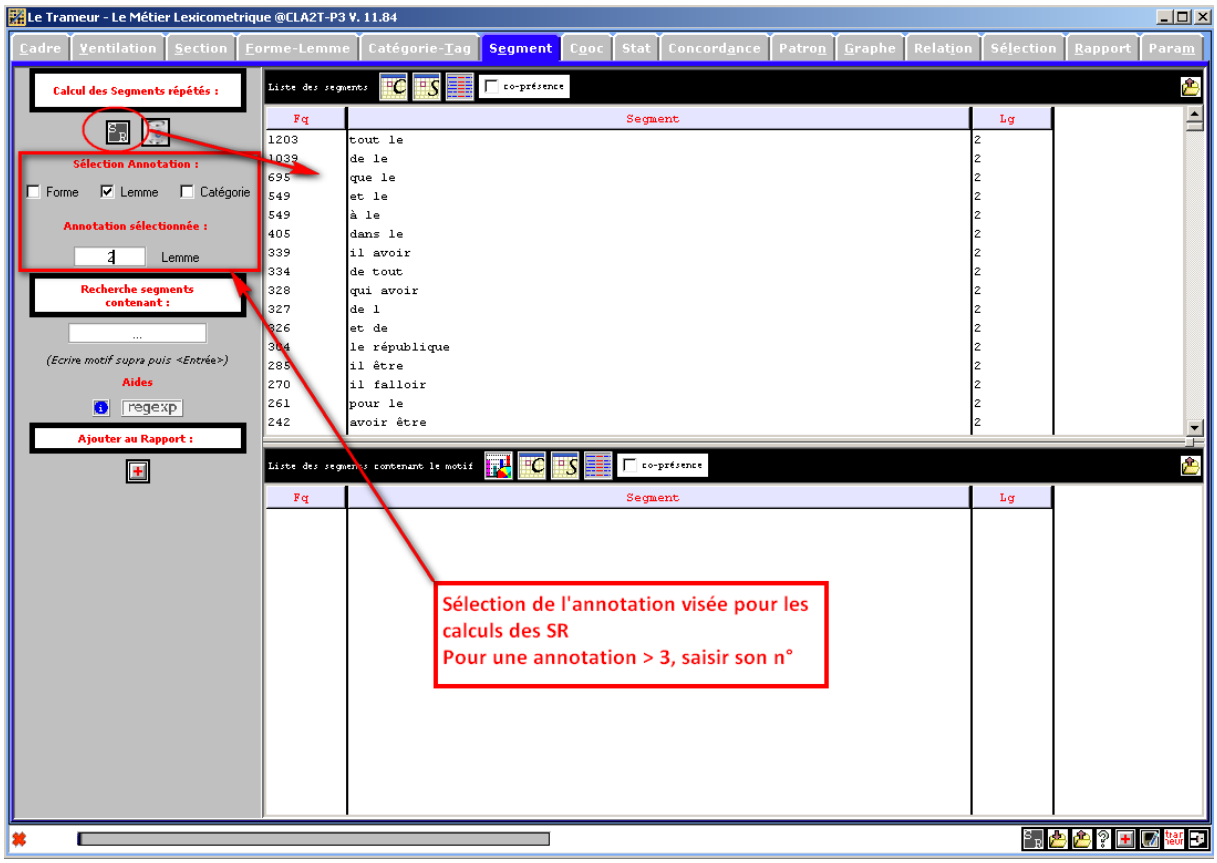


Figure 74 : ***Le Trameur*** **Onglet SEGMENT**

Remarque : Les segments apparaissent dans le volet supérieur de l'onglet **SEGMENT**.

Après sélection d'un segment (ou de plusieurs), il est possible de déclencher :

- la recherche de cette forme (ou de plusieurs) dans la représentation cartographique du ***Cadre*** *infra* (*cf* onglet **CADRE**).
- la recherche de cette forme (ou de plusieurs) dans la représentation cartographique de la carte des sections *infra* (*cf* onglet **SECTION**).
- une concordance des segments sélectionnés.

On peut aussi rechercher des segments contenant une forme donnée en utilisant les zones de saisie disponibles sur la gauche de l'onglet **SEGMENT**.

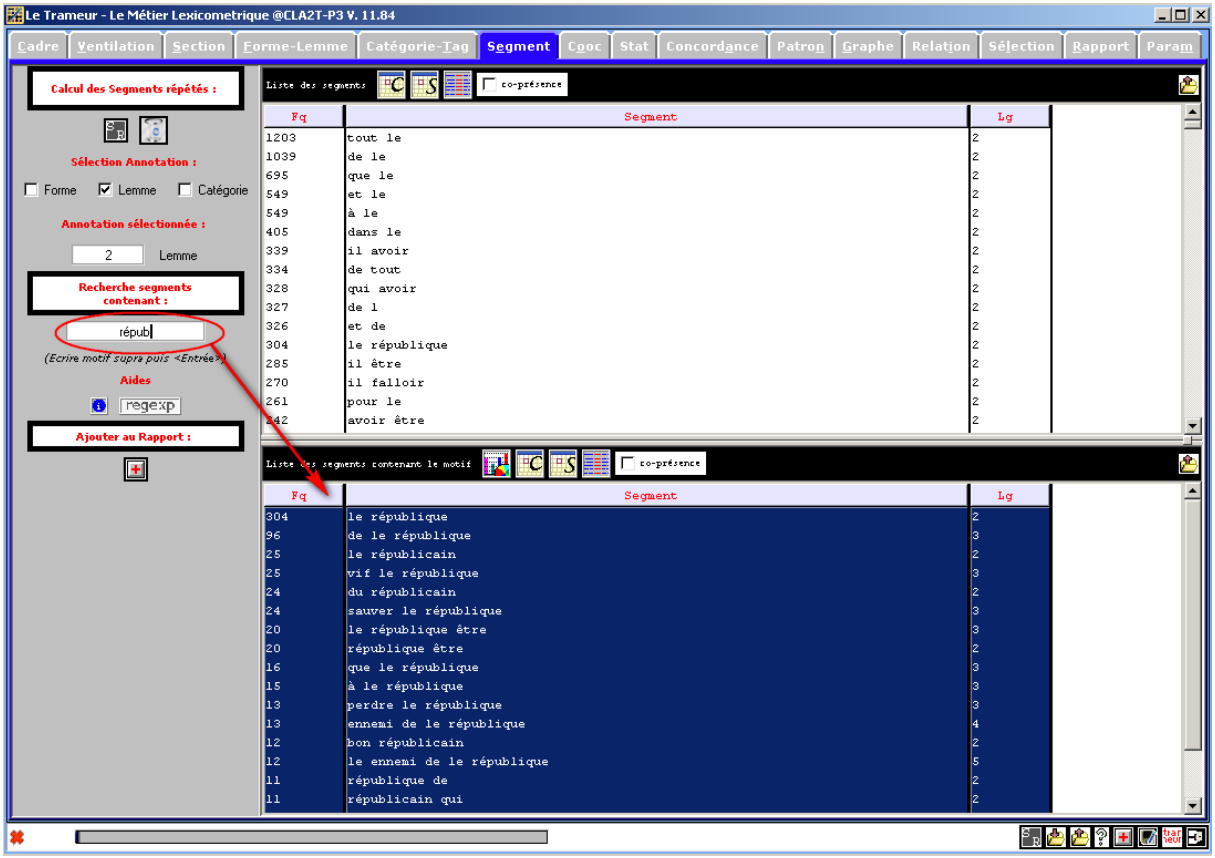


Figure 75 : *Le Trameur* Onglet SEGMENT

Après sélection d'un segment (ou de plusieurs), il est possible de déclencher :

- la recherche de ce segment (ou de plusieurs) dans la représentation cartographique du ***Cadre*** *infra* (*cf* onglet **CADRE**).
- la recherche de ce segment (ou de plusieurs) dans la représentation cartographique de la carte des sections *infra* (*cf* onglet **SECTION**).
- la concordance de ce segment (ou de plusieurs) (*cf* onglet **CONCORDANCE**).
- la ventilation de ce segment (ou de plusieurs) (*cf* onglet **VENTILATION**).

18 Cooccurrences – Poly-Cooccurrences

L'onglet **COOC** permet de lancer des modules de calculs de cooccurrences et de poly-cooccurrences sur la base chargée. Les modules disponibles s'inscrivent dans la démarche mise en œuvre dans le travail de William Martinez (2002, 2003, 2006).

- Une *cooccurrence* désigne l'apparition de deux mots en même temps et dans le même contexte.
- Le terme *poly-cooccurrence* désigne les attractions lexicales au-delà de la cooccurrence binaire.

Le module de poly-cooccurrences intégré reprend l'algorithme décrit dans [Martinez, 2006] :

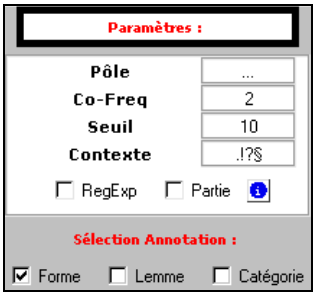
- On calcule pour le pôle A les cooccurrents spécifiques A_1, A_2 et A_3
  - Dans leurs contextes communs, on calcule pour les pôles $A+A_1$ les cooccurrents spécifiques A_{11} et A_{12}
  - Les pôles $A+A_1+A_{11}$ ont pour cooccurrent spécifique A_{111}
    - On examine les cooccurrents de $A+A_1+A_{11}+A_{111}$:
      - Les pôles $A+A_1+A_{11}+A_{111}$ n'ont pas de cooccurrent spécifique : l'exploration s'interrompt pour ce chemin
  - Les pôles $A+A_1+A_{12}$ ont pour cooccurrent spécifique A_{121}
    - On examine les cooccurrents de : $A+A_1+A_{11}+A_{121}$:
      - Etc.
  - Dans leurs contextes communs, on calcule pour les pôles $A+A_2$ les cooccurrents spécifiques A_{21} et A_{22}
    - Etc.
- Durant l'exploration, différents filtrages conditionnent l'épuisement des explorations contextuelles et réduisent le bruit dans les résultats pour privilégier l'information la plus spécifique : seuils maximaux de fréquence et de spécificité du cooccurrent.

18.1 Paramétrage et lancement des modules

Pour le calcul des cooccurrents, la zone de paramétrage de la forme pôle et des seuils permet de :

- Définir la forme pôle (simple ou complexe via une regexp)
- Définir une co-fréquence minimale
- Définir une spécificité minimale
- Définir une unité contextuelle pour le calcul
- Définir l'annotation à utiliser (forme, lemme, catégorie)

L'unité contextuelle est définie en fournissant soit une liste de caractères délimiteurs permettant de définir des contextes délimités par ces caractères, soit en spécifiant une zone du ***Cadre*** à utiliser pour délimiter les zones visées.



Dans la zone de paramétrage du calcul des cooccurrents, une case à cocher permet de définir le pôle sous la forme d'une expression régulière. Le calcul des cooccurrents/polycooccurrents peut se faire en selectionnant le niveau d'annotation[14] à prendre en compte pour construire ces séquences ; pour le moment, on peut selectionner : **Annotation n°1** (forme), **Annotation n°2** (lemme) ou **Annotation n°3** (categorie).

18.2 Paramétrage de l'unité de contexte

Dans ce qui suit, on considère que l'on travaille sur le texte suivant :

```xml
<?xml version="1.0" encoding="iso-8859-1"?>
<?xml-stylesheet type="text/xsl" href="q2.xsl"?>
<TEI>
    <teiHeader>
        <fileDesc>
            <titleStmt>
                <title>Poésies</title>
                <author>Arthur Rimbaud</author>
            </titleStmt>
        </fileDesc>
    </teiHeader>
    <text>
        <group>
            <text>
                <front>
                    <head>Le Dormeur du Val</head>
                    <dateline>7 octobre 1870</dateline>
                </front>
                <body>
                    <lg>
                        <lg>
                            <l>C'est un trou de verdure où chante une rivière</l>
                            <l>Accrochant follement aux herbes des haillons</l>
                            <l>D'argent ; où le soleil, de la montagne fière,</l>
                            <l>Luit : c'est un petit val qui mousse de rayons.</l>
                        </lg>
                        <lg>
                            <l>Un soldat jeune, bouche ouverte, tête nue,</l>
                            <l>Et la nuque baignant dans le frais cresson bleu,</l>
                            <l>Dort ; il est étendu dans l'herbe, sous la nue,</l>
                            <l>Pâle dans son lit vert où la lumière pleut.</l>
                        </lg>
                        <lg>
                            <l>Les pieds dans les glaïeuls, il dort. Souriant comme</l>
                            <l>Sourirait un enfant malade, il fait un somme :</l>
                            <l>Nature, berce-le chaudement : il a froid.</l>
                        </lg>
                        <lg>
                            <l>Les parfums ne font pas frissonner sa narine ;</l>
                            <l>Il dort dans le soleil, la main sur sa poitrine</l>
                            <l>Tranquille. Il a deux trous rouges au côté droit.</l>
                        </lg>
                    </lg>
                </body>
            </text>
        </group>
    </text>
</TEI>
```

18.2.1 Contexte défini par des délimiteurs

Pour le calcul des cooccurrents, la zone de paramétrage de la forme pôle et des seuils permet de :
L'unité contextuelle peut-être définie en fournissant une liste de caractères délimiteurs permettant de définir des contextes délimités par ces caractères. Par exemple, dans l'exemple précédent, si la liste des caractères définissant le contexte du calcul est réduite au caractère point (.), les contextes retenus seront :

[14] Illustration ici sur le corpus de Test "Base Prematures 96" : http://tal.univ-paris3.fr/trameur/base-trameur-p96annot_SR/index_07_6_2009_17_18_42.xml

Contexte n°1 :
```
Le dormeur du val
Poésies
Arthur Rimbaud
Le Dormeur du Val
7 octobre 1870
C'est un trou de verdure où chante une rivière
Accrochant follement aux herbes des haillons
D'argent ; où le soleil, de la montagne fière,
Luit : c'est un petit val qui mousse de rayons.
```

Contexte n°2 :
```
Un soldat jeune, bouche ouverte, tête nue,
Et la nuque baignant dans le frais cresson bleu,
Dort ; il est étendu dans l'herbe, sous la nue,
Pâle dans son lit vert où la lumière pleut.
```

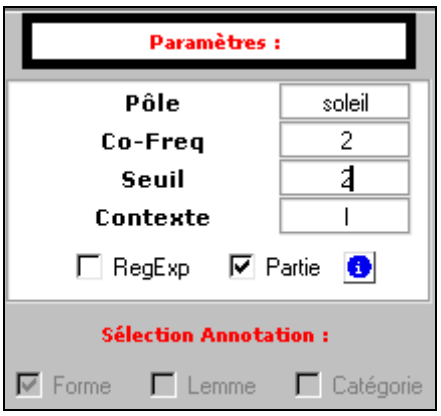
Contexte n°3 :
```
Les pieds dans les glaïeuls, il dort. Souriant comme
Sourirait un enfant malade, il fait un somme :
Nature, berce-le chaudement : il a froid.
```

Contexte n°4:
```
Les parfums ne font pas frissonner sa narine ;
Il dort dans le soleil, la main sur sa poitrine
Tranquille. Il a deux trous rouges au côté droit.
```

Au final, seuls les contextes contenant le pôle seront utilisés pour le calcul des cooccurrents.

18.2.1 Contexte défini par le *Cadre*

Il est aussi possible de considérer que les contextes à conserver pour le calcul correspondent à des zones induites par le ***Cadre***. Dans notre exemple, si on souhaite retenir que les contextes de type « vers », il suffit d'indiquer que les contextes à conserver sont ceux définis par la balise `l`. La mise en œuvre dans le ***Trameur*** se fait de la manière suivante :



Le paramétrage retenu ci-dessus est donc :

- la forme pôle : `soleil`
- co-fréquence minimale : `2`
- spécificité minimale : `2`
- unité contextuelle pour le calcul : la balise `l` (case à cocher « Partie » active)

- annotation à utiliser : `forme`

La liste des contextes retenus dans un premier temps rassemble donc l'ensemble des vers :

```
C'est un trou de verdure où chante une rivière
Accrochant follement aux herbes des haillons
D'argent ; où le soleil, de la montagne fière,
Luit : c'est un petit val qui mousse de rayons.
Un soldat jeune, bouche ouverte, tête nue,
Et la nuque baignant dans le frais cresson bleu,
Dort ; il est étendu dans l'herbe, sous la nue,
Pâle dans son lit vert où la lumière pleut.
Les pieds dans les glaïeuls, il dort. Souriant comme
Sourirait un enfant malade, il fait un somme :
Nature, berce-le chaudement : il a froid.
Les parfums ne font pas frissonner sa narine ;
Il dort dans le soleil, la main sur sa poitrine
Tranquille. Il a deux trous rouges au côté droit.
```

Et au final, les contextes retenus par le calcul sont parmi les précédents vers ceux contenant le pôle :

```
D'argent ; où le soleil, de la montagne fière,
Il dort dans le soleil, la main sur sa poitrine
```

18.3 Mise en œuvre du calcul

Les figures qui suivent illustrent la démarche mise en œuvre pour construire des listes de cooccurrents puis des graphes donnant à voir les poly-cooccurrents.

Utilisation d'une stop-liste

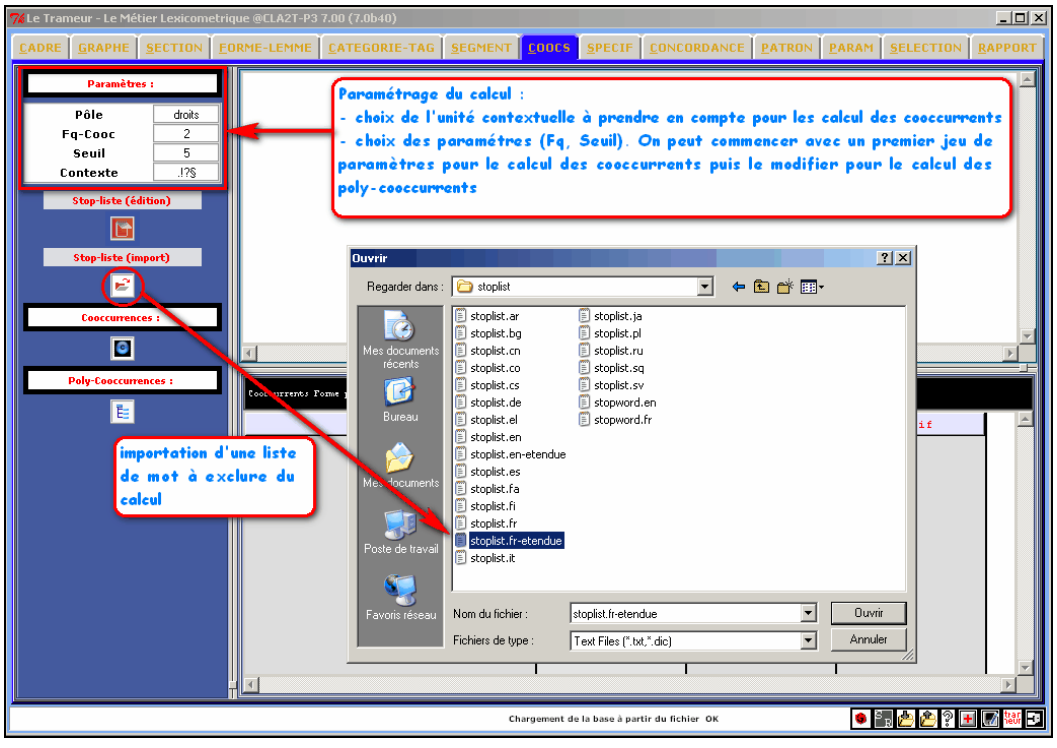


Figure 76 : *Le Trameur* Onglet COOC, importation d'une stop-liste

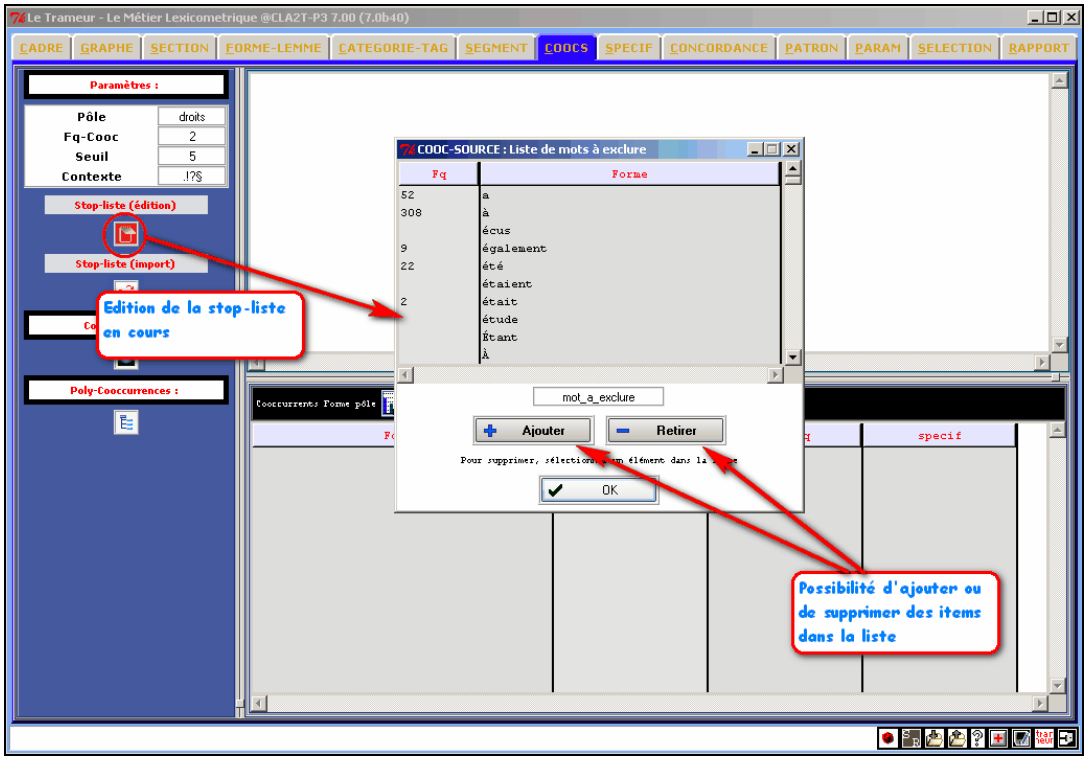
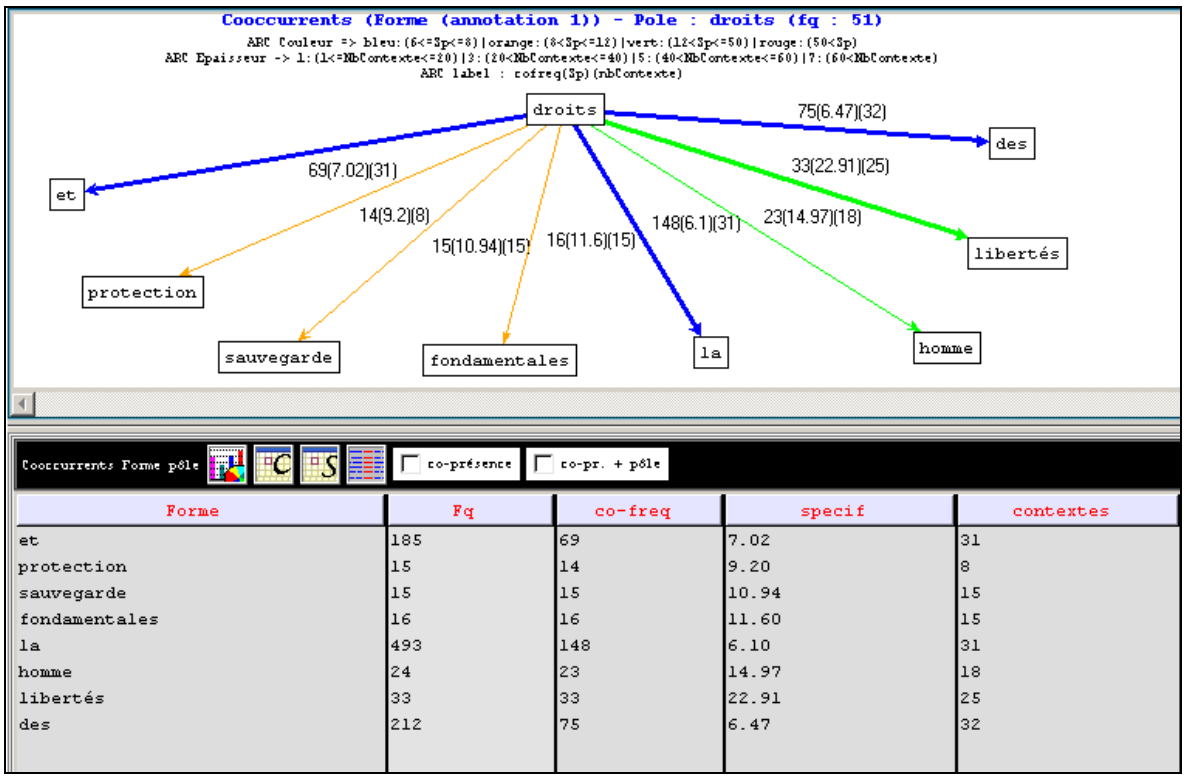


Figure 77 : *Le Trameur* Onglet COOC, édition de la stop-liste

Etape n°1 : Calcul des cooccurrents



Forme	Fq	co-freq	specif	contextes
et	185	69	7.02	31
protection	15	14	9.20	8
sauvegarde	15	15	10.94	15
fondamentales	16	16	11.60	15
la	493	148	6.10	31
homme	24	23	14.97	18
libertés	33	33	22.91	25
des	212	75	6.47	32

Figure 78 : *Le Trameur* Onglet COOC, étape n°1 : module Cooccurrents

Etape n°2 : Calcul des polycooccurrents

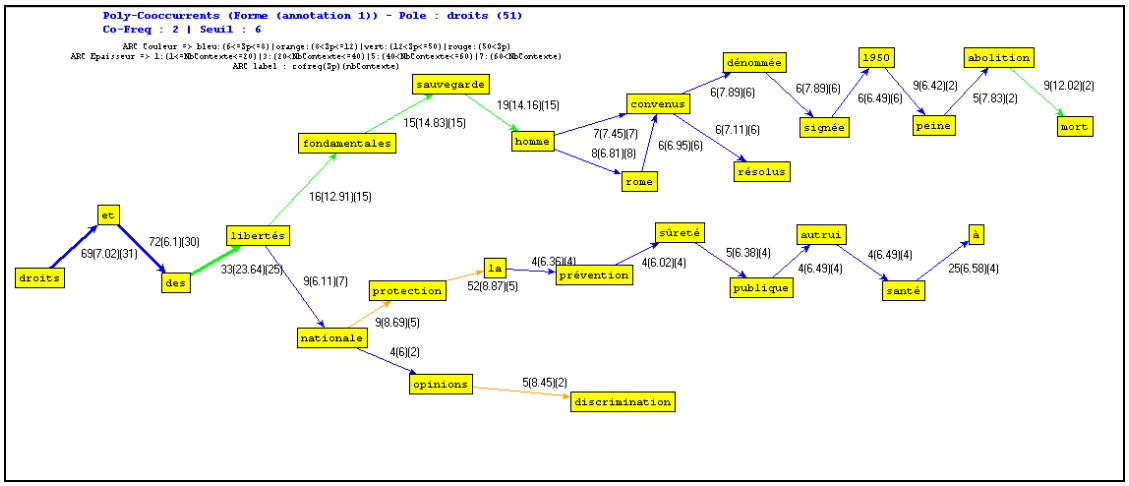


Figure 79 : *Le Trameur* Onglet COOC, étape n°2 : module Polycooccurrents

Dans les 2 graphes produits, les arcs connectant des nœuds portent un label regroupant 3 indications (disponibles aussi dans les colonnes de la liste des cooccurrents) : *co-fréquence (spécificité) (contextes)*. Par exemple, pour la forme *et* à partir du pôle *droits* on trouve la notation `69(7.02)31` qui indique que le cooccurrent apparaît 69 fois à côté du pôle avec une spécificité de 7.02 dans un total de 31 contextes. La représentation de ces arcs permet aussi de mettre en avant de manière graphique 2 de ces indications statistiques : la couleur du trait dépend de l'indice de spécificité et l'épaisseur du trait dépend du nombre de contextes (la légende précise les valeurs dans les 2 cas).

18.4 Cooccurrents et segments répétés

Le calcul des cooccurrents / polycooccurrents peut aussi être activé en intégrant les segments répétés si ce module a déjà été lancé.

Choix d'un pôle de type SR

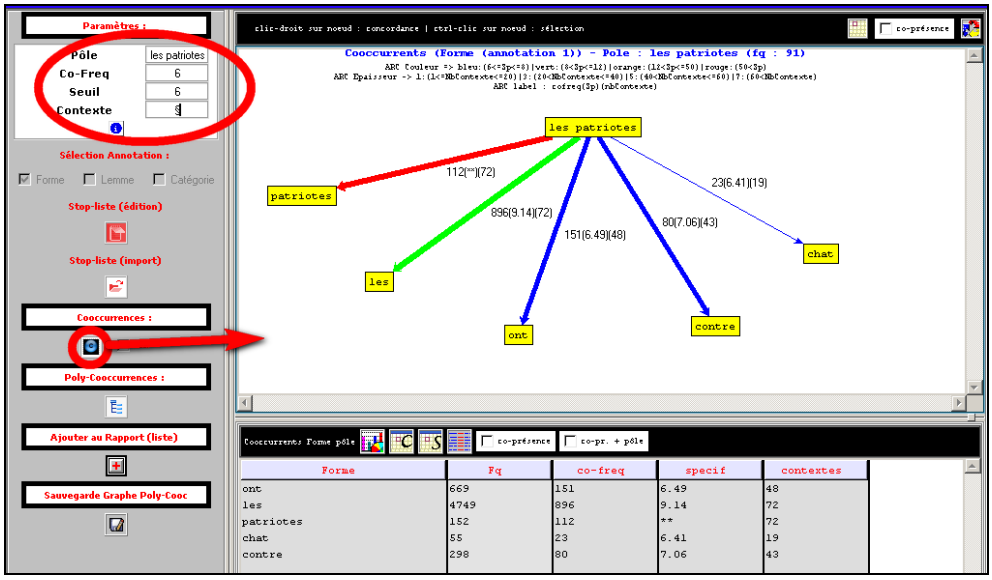


Figure 80 : *Le Trameur* Onglet COOC, un pôle de type SR

Si le calcul des SR a déjà été activé, il est possible de calculer les cooccurrents / polycooccurrents d'un pôle exprimé sous la forme d'un segment répété.

Intégration des SR dans les calculs de cooccurrence

De même, il est possible d'intégrer les segments disponibles dans le calcul des cooccurrents / polycooccurrents : la case à cocher idoine rend cette option disponible.

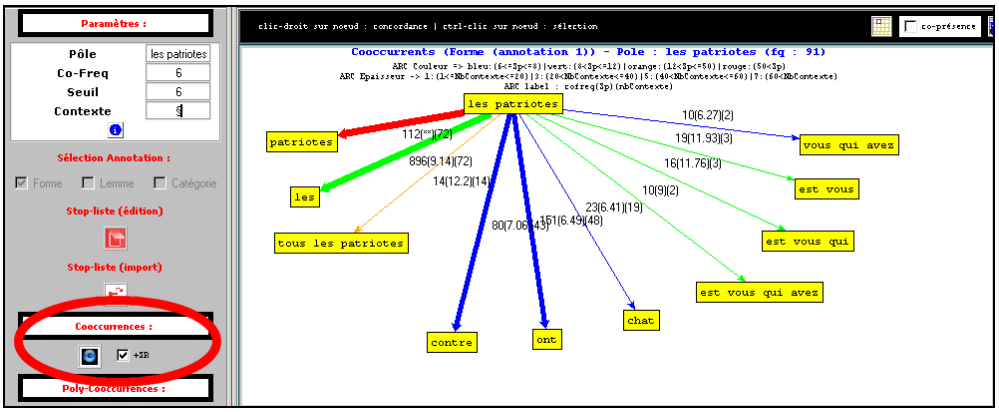
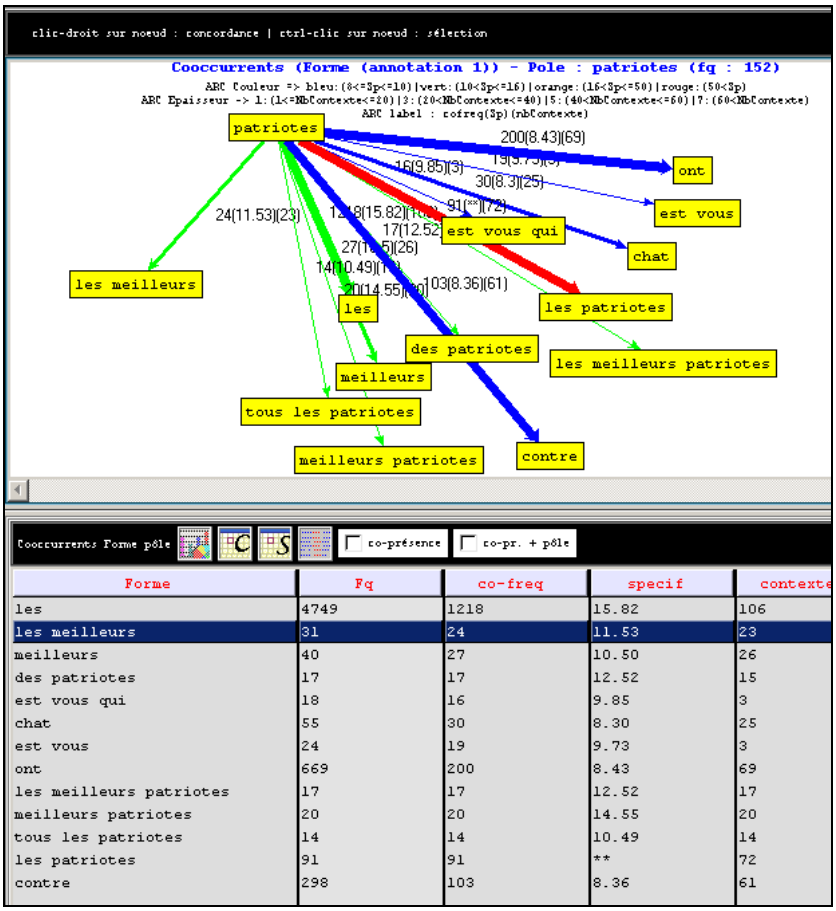


Figure 81 : *Le Trameur* Onglet COOC, Coocs + SR

18.5 Sélection des chemins de cooccurrences à examiner



Forme	Fq	co-freq	specif	contexte
les	4749	1218	15.82	106
les meilleurs	31	24	11.53	23
meilleurs	40	27	10.50	26
des patriotes	17	17	12.52	15
est vous qui	18	16	9.85	3
chat	55	30	8.30	25
est vous	24	19	9.73	3
ont	669	200	8.43	69
les meilleurs patriotes	17	17	12.52	17
meilleurs patriotes	20	20	14.55	20
tous les patriotes	14	14	10.49	14
les patriotes	91	91	**	72
contre	298	103	8.36	61

Figure 82 : *Le Trameur* Onglet COOC, Filtrage des polycoocs

Le calcul des chemins de polycoocurrence se fait sur la base de la liste calculée des cooccurrents du pôle choisi.

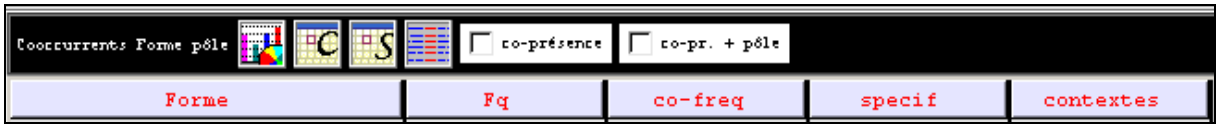
- Par défaut le calcul des polycooccurrents est réalisé en traitant tous les cooccurrents du pôle.
- Il est aussi possible de lancer le calcul des polycooccurrents sur une sélection de cooccurrents du pôle.

Dans la figure ci-dessus, la sélection d'un seul cooccurrent du pôle choisi permettra de ne calculer que la polycooccurrence initiée par cette sélection (coprésence du pôle et du cooccurrent sélectionnée).

18.6 Fonctionnalités complémentaires sur les résultats produits

Les résultats produits par ces 2 modules sont insérés dans l'onglet **COOC** et sont associés à des processus de calcul déjà disponibles dans d'autres onglets :

a. Actions sur la liste des cooccurrents :



Il suffit de sélectionner un ou plusieurs éléments dans la liste des cooccurrents puis d'activer l'opération visée (*ventilation, ventilation sur* ***Cadre****, ventilation sur carte des sections, concordance*).

b. Actions sur le graphe des cooccurrent ou des polycoccurrents

- le raccourci *clic-droit* sur un nœud du graphe déclenche le calcul de la concordance de la forme associée au nœud visé
- le raccourci *control-clic* sur un nœud permet de le sélectionner

Des nœuds du graphe étant sélectionnés, on peut ensuite les projeter sur la carte des sections :

Si l'option « co-présence » est activée, les sections mises au jour seront celles contenant l'ensemble des formes visées (ET logique) : on peut ainsi visualiser en contexte les chemins de polycooccurrences calculés précédemment.

19 Extraction de patrons et graphes de mots

19.1 Extraction de patrons (recherche sur un niveau d'annotation)

Une base de travail étant chargée dans le ***Trameur***, il est possible de réaliser une extraction de patron sur la base des annotations associées à chaque item de la ***Trame***. Les séquences de termes à construire pour un patron donné peuvent être définies sur les catégories morphosyntaxiques déterminées par *treetagger* ou sur les annotations importées. Pour lancer cette recherche de patrons, il faut se placer dans l'onglet **PATRON** :

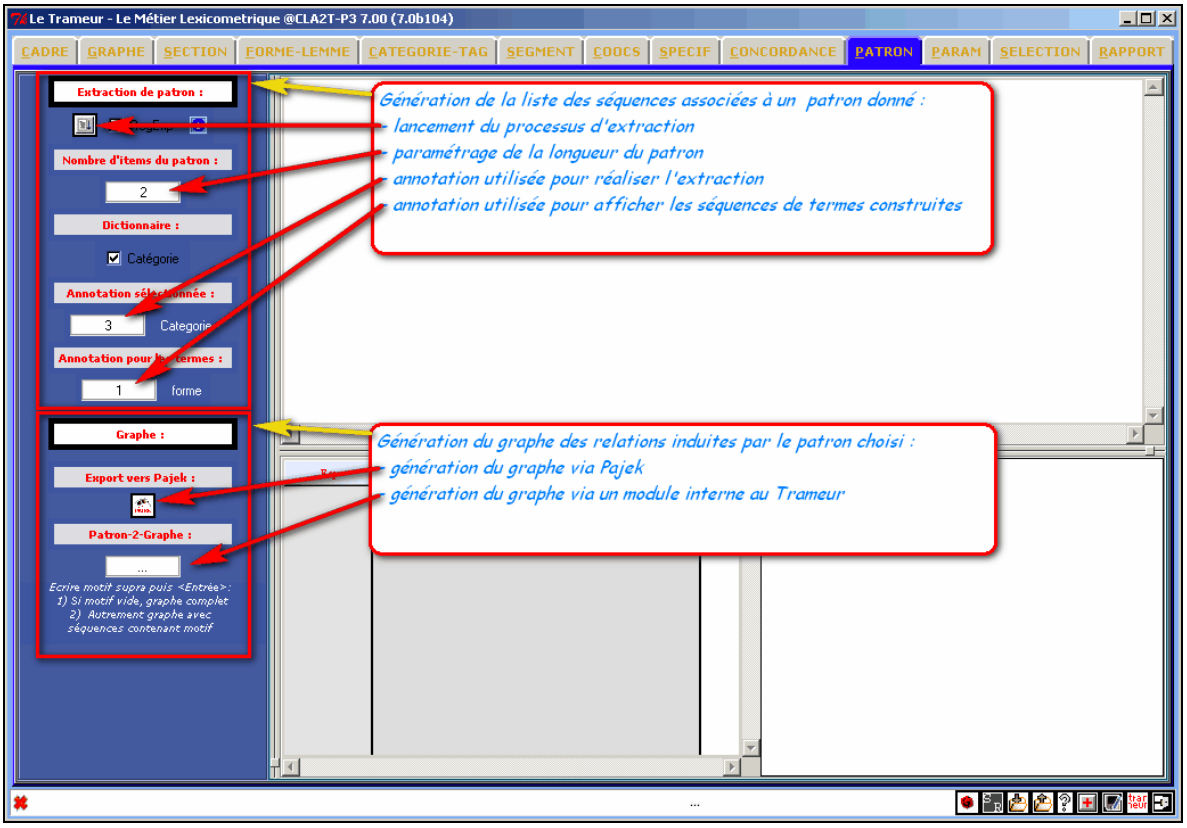


Figure 83 : *Le Trameur* onglet PATRON

Dans les figures qui suivent, nous allons illustrer cette extraction de patron en utilisant notre fichier de travail et en recherchant tout d'abord le patron « `NOM ADJ` » sur l'annotation n°3 (les catégories construites par *treetagger*) puis le patron « `orange rouge bleu` » sur l'annotation n°4 importée, les couleurs associées à chaque item de la ***Trame*** (*cf* partie suivante pour une présentation de l'importation d'annotation). Avant de lancer la génération de la liste de séquences de termes, il convient de choisir le niveau d'annotation choisi pour définir le patron ; par défaut, ce niveau (le numéro 3) est celui qui est associé aux catégories construites par *treetagger*. Pour choisir un numéro correspondant à un système d'annotation importé, il suffit de donner son numéro (numéro 4[15] ou supérieur). Pour enregistrer le patron visé, il faut activer le bouton . Ce bouton déclenche l'apparition d'une fenêtre contenant un certain nombre de zones de saisie permettant de définir les constituants du patron. Le nombre de zones de saisie correspond à la valeur (modifiable) associée à l'étiquette :

[15] Rappel : par défaut, les numéros d'annotation 1, 2 et 3 sont associés respectivement aux formes, aux lemmes et aux catégories (que *treetagger* soit activé ou non).

« **Nombre d'items du patron** » dans l'onglet **PATRON**. L'écriture des constituants d'un patron peut se faire en utilisant des expressions régulières si ce paramétrage a été au préalable défini dans l'onglet **PATRON**.

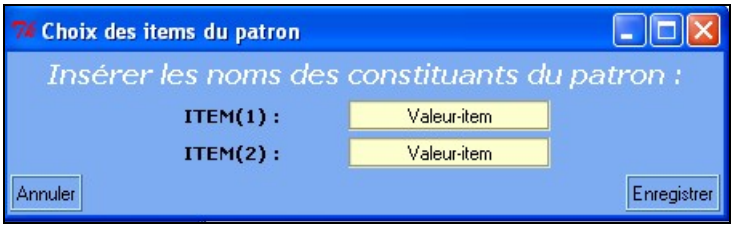


Figure 84 : *Le Trameur* zone d'enregistrement d'un patron

Après avoir renseigné les valeurs des éléments constitutifs du patron visé, le bouton « *Enregistrer* » déclenche la génération des séquences de termes (*via* leurs formes) associées au patron visé. Cette liste sera intégrée dans la zone d'édition gauche de l'onglet **PATRON**.

Exemple 1 :

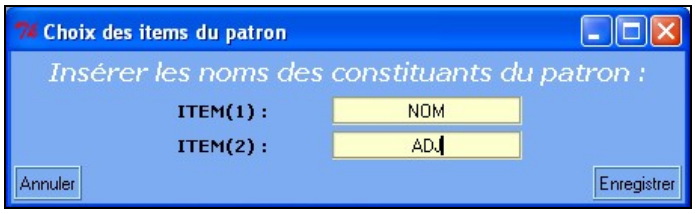


Figure 85 : *Le Trameur* écriture d'un patron

Figure 86 : *Le Trameur* liste des séquences associées au patron

Exemple 2 :

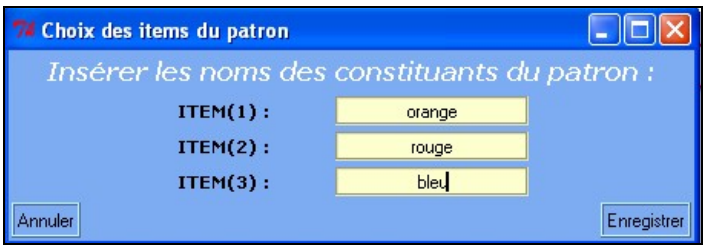


Figure 87 : *Le Trameur* écriture et paramétrage d'un patron

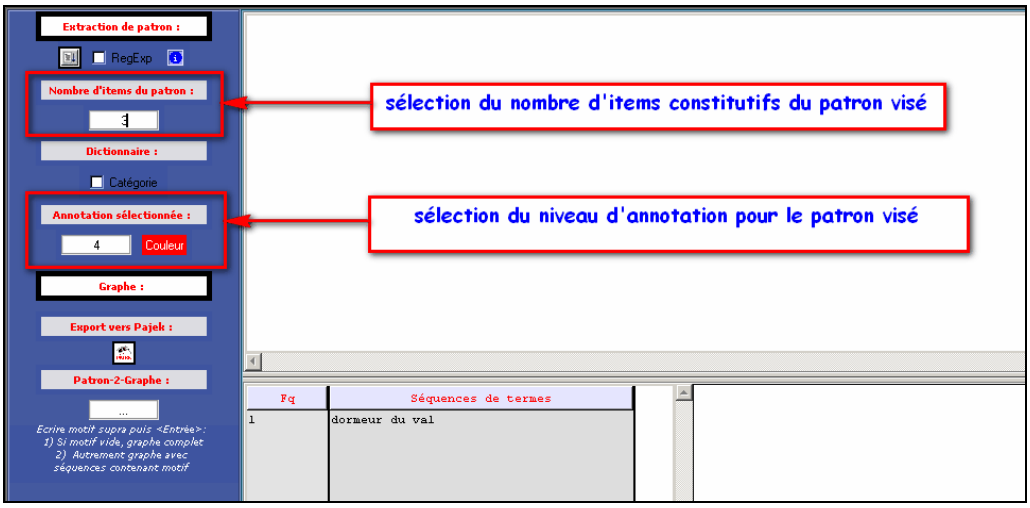


Figure 88 : *Le Trameur* liste des séquences associées au patron

dans notre texte de travail, les couleurs associés aux formes sont, pour le début du texte :

```
Le-vert dormeur-orange  du-rouge val-bleu C-vert '-jaune
```

19.2 Extraction de patrons (avec croisement d'annotations)

Cette fonctionnalité permet d'extraire un patron en décrivant initialement ces composants sur des niveaux d'annotations différents. Pour illustrer cette fonctionnalité, nous allons utiliser la base Prématurés 96[16] : dans cette base, chaque item de la ***Trame*** est associé à 4 annotations : forme (niveau 1), lemme (niveau 2), catégorie (niveau 3), annotation sémantique (niveau 4).

1. *Paramétrage de l'extraction :*

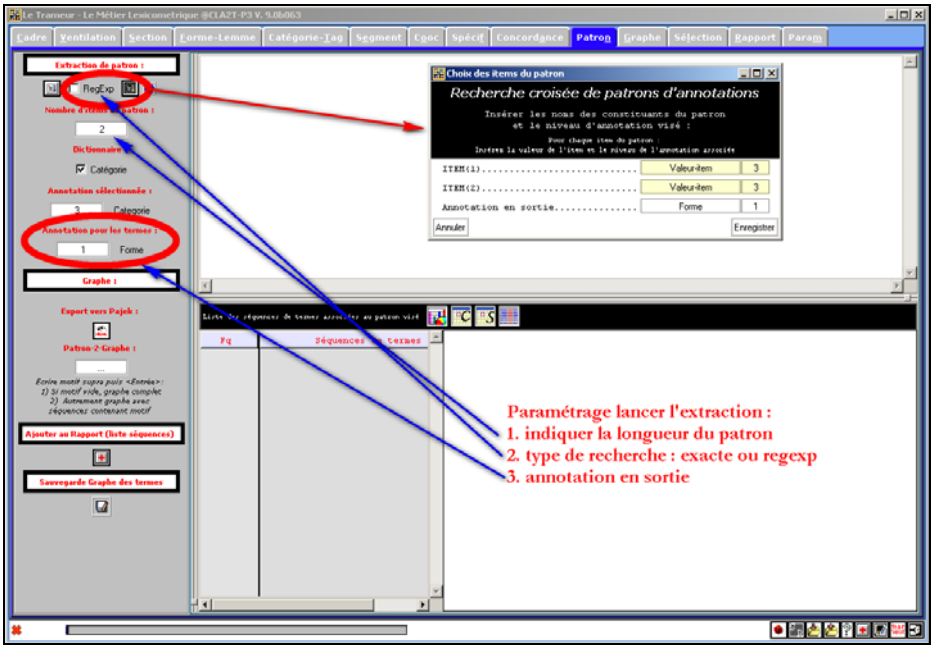


Figure 89 : *Le Trameur* Extraction de patron (croisement d'annotation)

[16] http://tal.univ-paris3.fr/trameur/bases/base-p96-trameur.zip

2. *Paramétrage du patron*

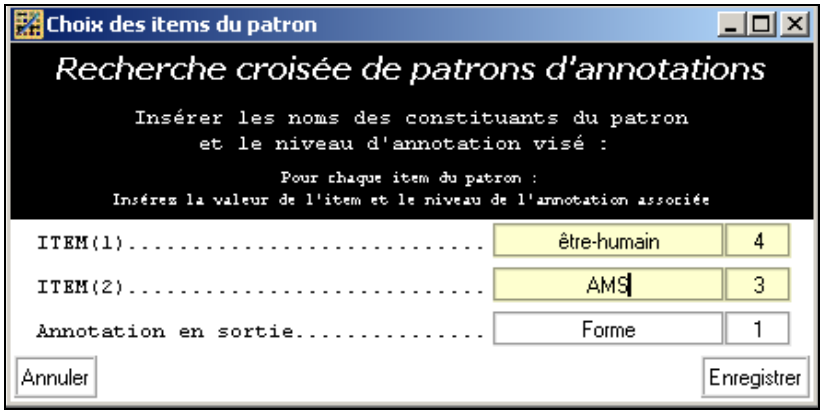
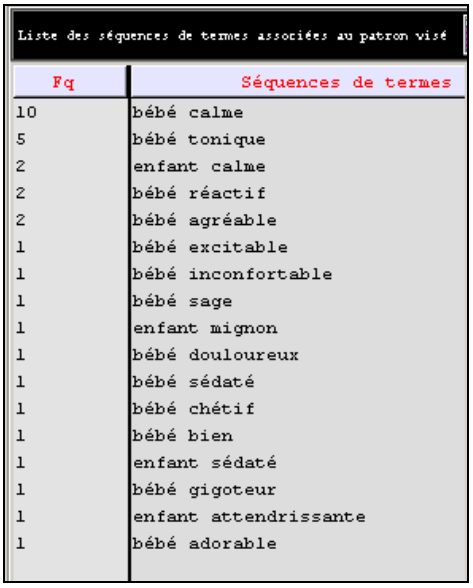

Choix des items du patron
Recherche croisée de patrons d'annotations
Insérer les noms des constituants du patron
et le niveau d'annotation visé :
Pour chaque item du patron :
Insérez la valeur de l'item et le niveau de l'annotation associée
ITEM(1)..... être-humain 4
ITEM(2)..... AMS 3
Annotation en sortie..... Forme 1
Annuler
Enregistrer

Figure 90 : *Le Trameur* Extraction de patron (croisement d'annotation)

On cherche donc ici un patron de longueur 2, dont le premier composant porte la valeur « être-humain » (annotation de niveau 4 : sémantique) et dont le second porte la valeur AMS (Adjectif Masculin Singulier) (annotation de niveau 3 : POS). Les termes extraits seront affichés suivant leur forme graphique.

Résultat produit : patron « être-humain AMS »



Fq	Séquences de termes
10	bébé calme
5	bébé tonique
2	enfant calme
2	bébé réactif
2	bébé agréable
1	bébé excitable
1	bébé inconfortable
1	bébé sage
1	enfant mignon
1	bébé douloureux
1	bébé sédaté
1	bébé chétif
1	bébé bien
1	enfant sédaté
1	bébé gigoteur
1	enfant attendrissante
1	bébé adorable

Figure 91 : *Le Trameur* Extraction de patron (croisement d'annotation)

Si la case à cocher *RegExp* est activée, l'écriture des composants du patron peut se faire en décrivant ces composants par une expression régulière.

19.3 Actions sur les séquences extraites à partir d'un patron donné

Les séquences de termes associées à un patron donné sont réutilisables pour des opérations déjà disponibles dans les autres onglets :

- *Concordance* : calcul d'une concordance sur toutes les positions des séquences sélectionnées
- *Ventilation-section* : ventilation sur la carte des sections de toutes les séquences sélectionnées

- *Ventilation-**Cadre*** : ventilation sur le ***Cadre*** de toutes les séquences sélectionnées
- *Ventilation-partie* : ventilation de toutes les séquences sélectionnées sur une partie donnée

Les figures qui suivent illustrent de nouveau la démarche à suivre pour lancer une recherche de patron puis pour déclencher les opérations sur des éléments de la liste de termes produite :

a. Paramétrage de l'extraction de patron

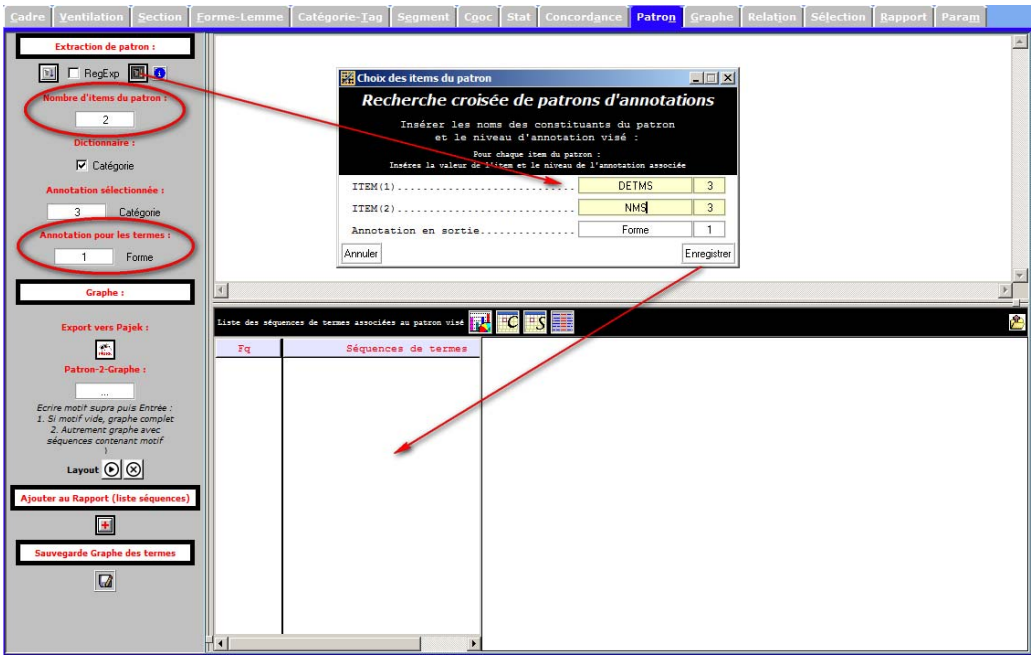


Figure 92 : *Le Trameur* Paramétrage de l'extraction du patron visé et de l'affichage des séquences extraites

b. Sélection des termes puis activation des calculs visés :

Figure 93 : *Le Trameur* Affichage des séquences extraites sur la base du patron visé

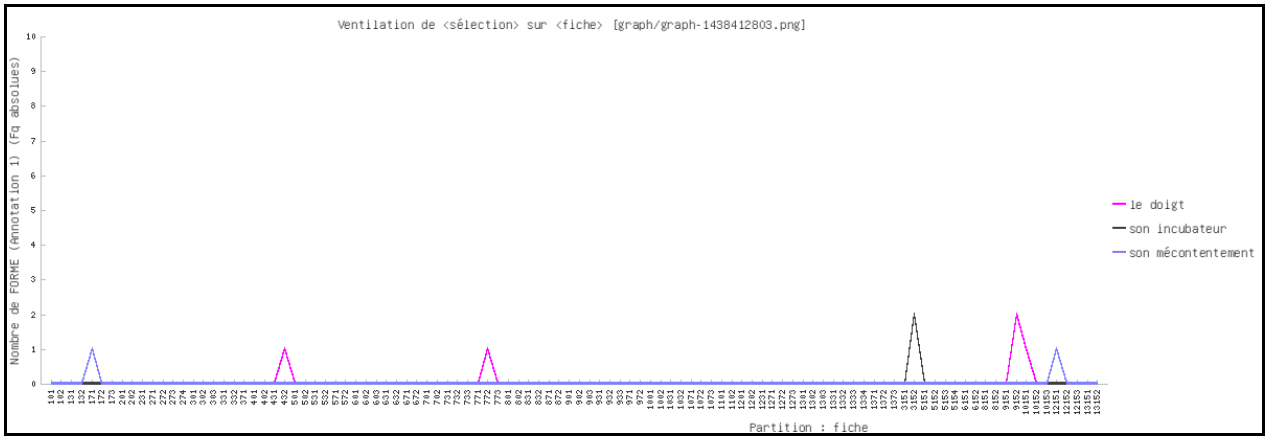


Figure 94 : *Le Trameur* Ventilation-partie d'une liste de séquences sélectionnées

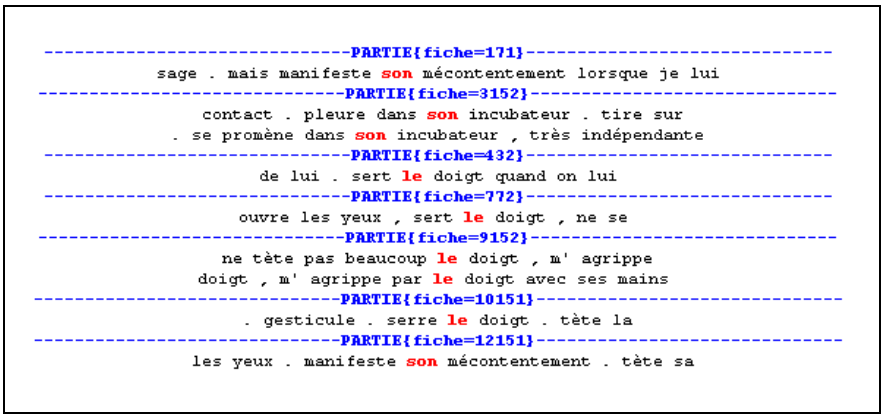

----------PARTIE{fiche=171}----------
sage . mais manifeste son mécontentement lorsque je lui
----------PARTIE{fiche=3152}----------
contact . pleure dans son incubateur . tire sur
. se promène dans son incubateur , très indépendante
----------PARTIE{fiche=432}----------
de lui . sert le doigt quand on lui
----------PARTIE{fiche=772}----------
ouvre les yeux , sert le doigt , ne se
----------PARTIE{fiche=9152}----------
ne tète pas beaucoup le doigt , m' agrippe
doigt , m' agrippe par le doigt avec ses mains
----------PARTIE{fiche=10151}----------
. gesticule . serre le doigt . tète la
----------PARTIE{fiche=12151}----------
les yeux . manifeste son mécontentement . tète sa

Figure 95 : *Le Trameur* Concordance d'une liste de séquences sélectionnées

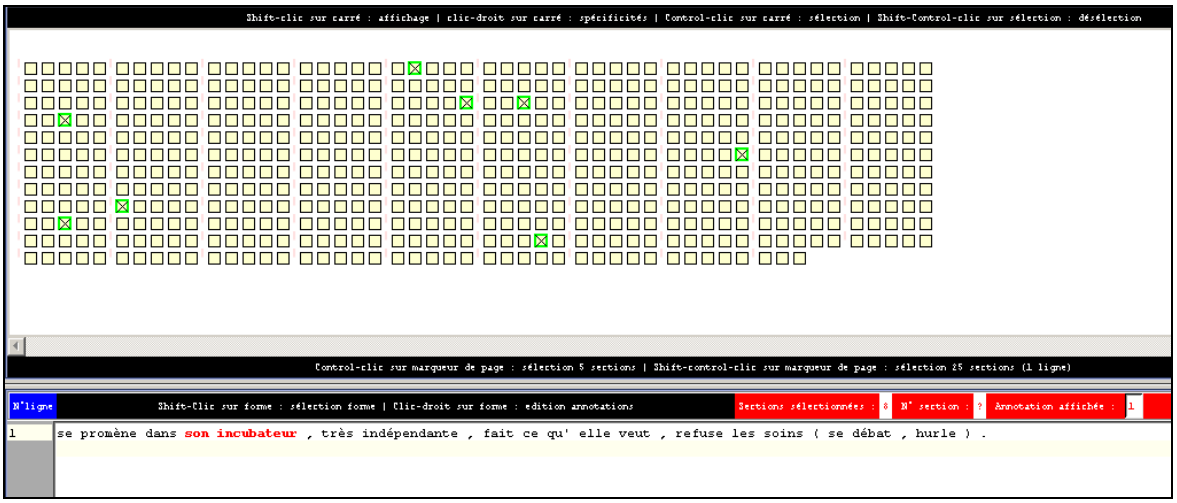

Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : désélection
Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)
N°ligne
Shift-Clic sur forme : sélection forme | Clic-droit sur forme : edition annotations
Sections sélectionnées :
N° section :
Annotation affichée :
1
se promène dans son incubateur , très indépendante , fait ce qu' elle veut , refuse les soins (se débat , hurle) .

Figure 96 : *Le Trameur* Ventilation-section d'une liste de séquences sélectionnées

19.4 Graphes de mots avec Le Trameur

On peut ensuite représenter la liste de séquences de termes associés à un patron sous la forme d'un graphe de mots donnant à voir la proximité entre ces mots induite par le patron visé. Dans l'exemple présenté infra, le texte de travail est un extrait d'un débat télévisé.

Pour générer ce graphe, il convient de renseigner la zone de saisie « *Patron-2-graphe* » :

1. si le motif saisi est vide, le programme construit le graphe complet donnant à voir l'ensemble des relations entre les mots des séquences construites

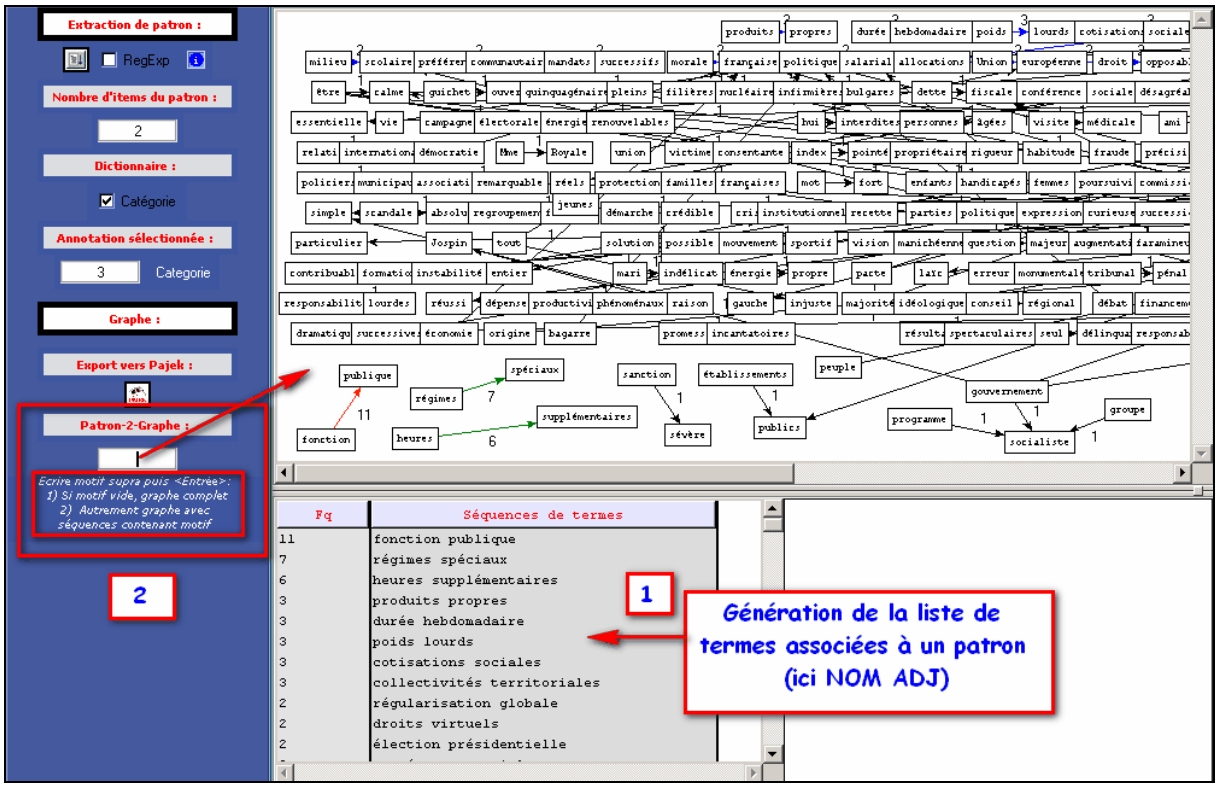


Figure 97 : *Le Trameur* génération du graphe de mots complet associé à un patron

2. si le motif saisi n'est pas vide, le programme construit le graphe donnant à voir l'ensemble des relations entre les mots des séquences construites et contenant le motif.

Dans l'exemple ci-dessous, le motif « public » permet de donner à voir toutes les relations entre les éléments des séquences construites contenant ce motif.

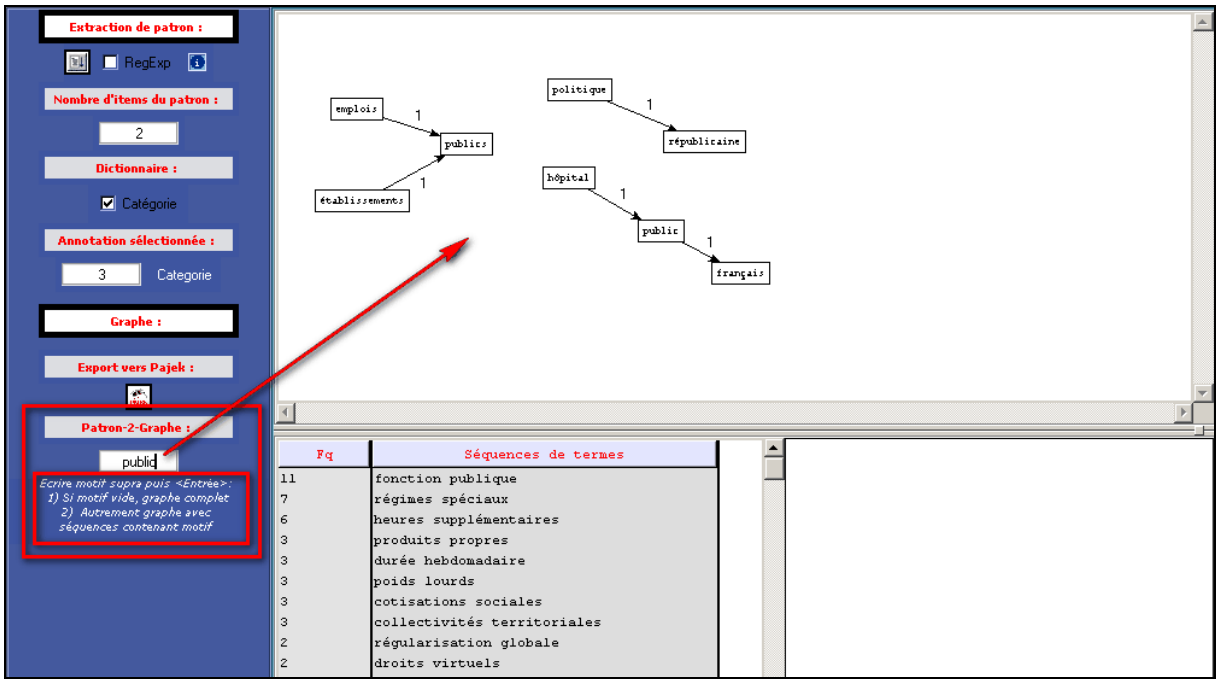


Figure 98 : *Le Trameur* génération du graphe de mots associé à un patron et contenant un motif

19.5 Graphes de mots avec Pajek

On peut aussi exporter la liste de séquences de termes associées à un patron dans un format compatible avec l'outil de représentation de réseaux *Pajek*[17]. On se reportera à la page en ligne « *Des Textes aux graphes* » pour avoir des informations complémentaires sur la démarche suivie.

Cet export se déroule en 2 temps :

1. Transformation du fichier contenant la liste des séquences associées à un patron au format `GRAPHML`[18].
2. Génération de cette même liste au format *Pajek* en utilisant une feuille de styles fournie avec le ***Trameur*** (`GraphML2Pajek.xsl` disponible dans le répertoire de travail du ***Trameur***).

Pour déclencher cet export, il faut activer le bouton . L'export terminé, le fichier au format `GRAPHML` est inséré dans la zone d'édition droite de l'onglet **PATRON** et une nouvelle fenêtre d'édition est activée, elle contient le fichier au format *Pajek*.

[17] *Hypertoire* : http://vlado.fmf.uni-lj.si/pub/networks/pajek/. *Mode d'emploi* : http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf

[18] http://graphml.graphdrawing.org/

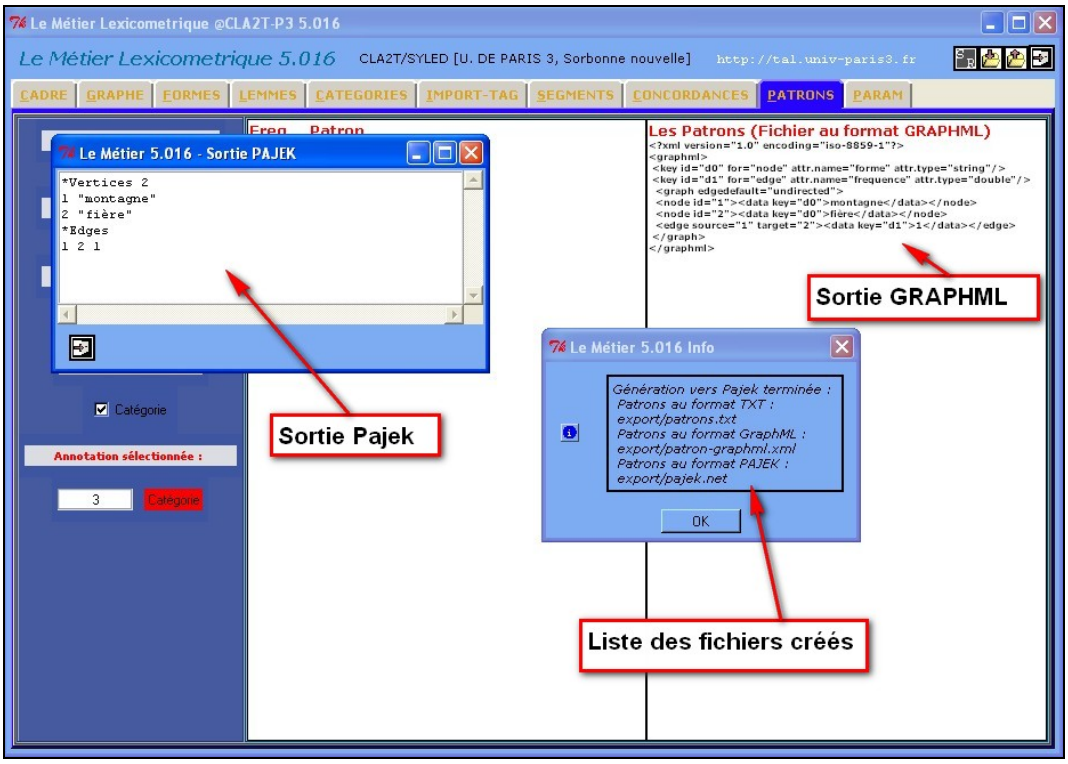


Figure 99 : *Le Trameur* export au format *Pajek*

La figure qui suit est une illustration d'un graphe construit avec *Pajek*. Le texte de travail est le texte `Duchn.txt` (*Le Père Duchesne*) disponible dans le dossier `textes`. Le patron visé est là encore `NOM ADJ`.

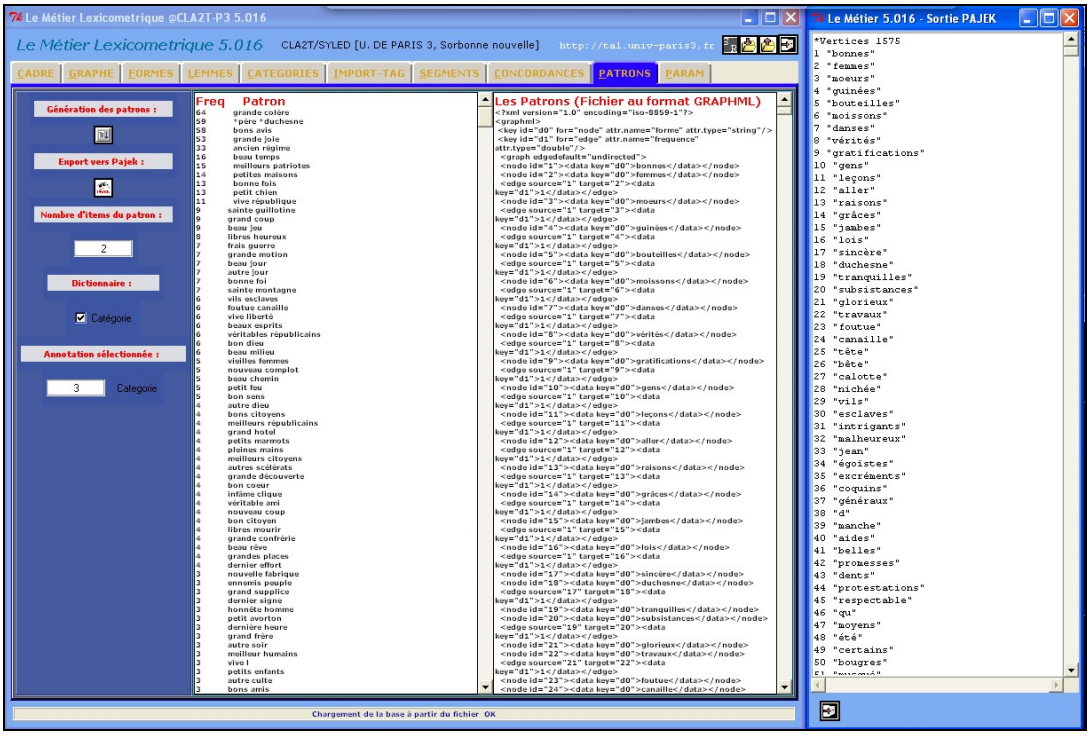


Figure 100 : *Le Trameur* export `NOM ADJ` au format *Pajek* sur Duchn

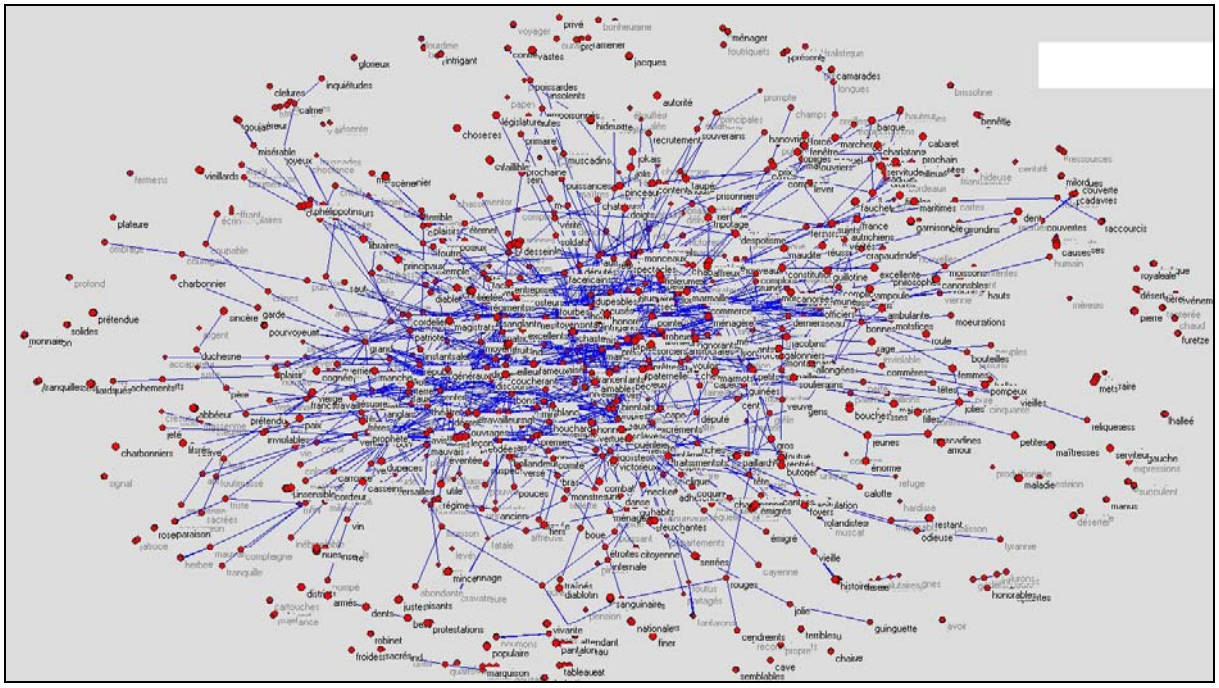


Figure 101 : Le réseau `NOM ADJ` de Duchn *via Pajek*

Dans Pajek, cliquer sur « *Draw* », puis sur « *Layout* » et choisir l'option, puis sur « *Export* » pour exporter le résultat.

20 Calculs textométriques complémentaires

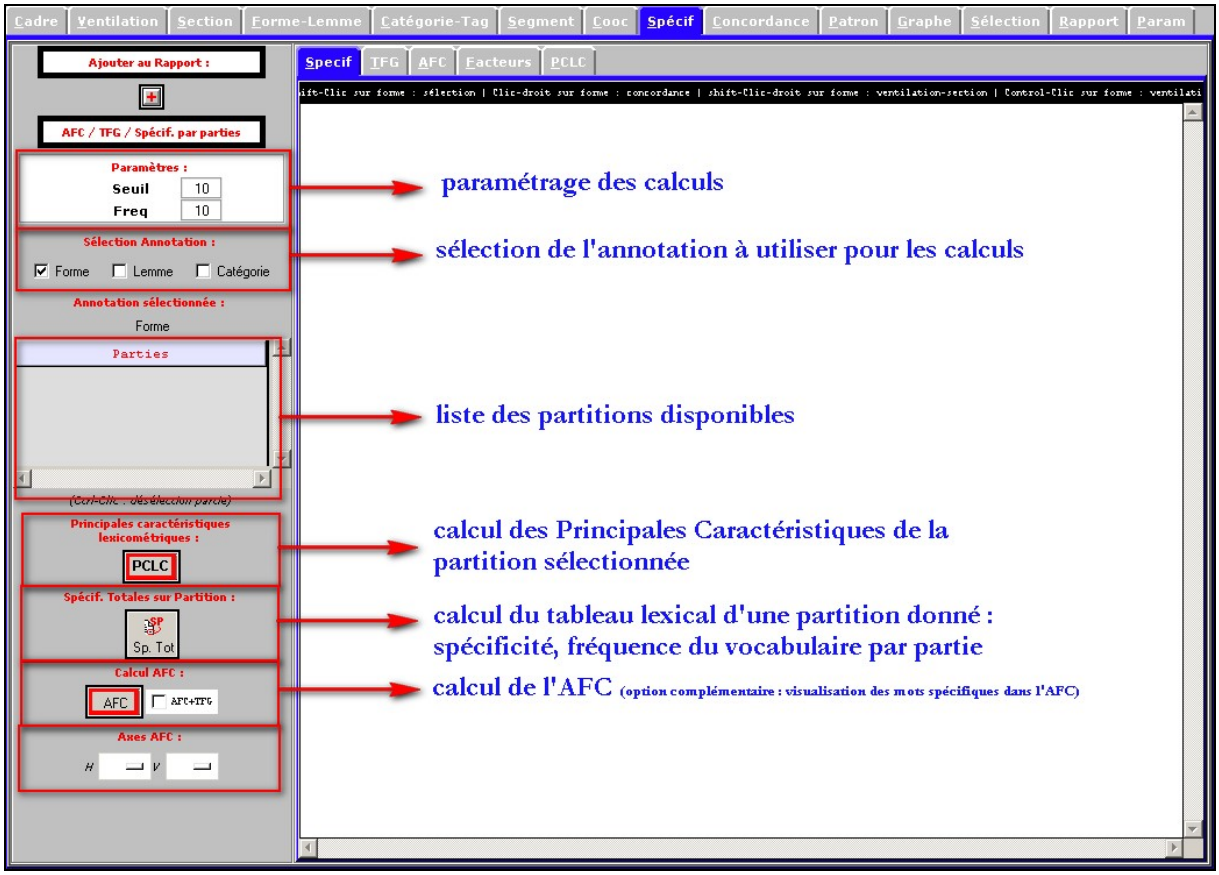


Figure 102 : Calculs textométriques complémentaires

Dans cette partie, nous décrivons les fonctionnalités disponibles dans l'onglet **SPECIF**. Ces différents calculs se font en sélectionnant au préalable le niveau d'annotation à prendre en compte dans les calculs ; pour le moment, on peut sélectionner : **Annotation n°1** (forme), **Annotation n°2** (lemme) ou **Annotation n°3** (catégorie).

Figure 103 : Sélection du niveau d'annotation pour calculs lexicométriques complémentaires

20.1 Tableau lexical d'une partition

Le bouton

permet de calculer le *tableau lexical* des items (forme ou lemme ou catégorie) d'une partition sélectionnée dont la fréquence est supérieure à un seuil donné et dont l'indice de spécificité est lui aussi supérieur à un seuil donné. Les lignes de ce tableau à double entrée sont constituées par les ventilations et les indices de spécificité des différents items dans les différentes parties de la partition choisie. Le terme générique *k(i,j)* du tableau

est égal au nombre de fois que l'item *i* est attestée dans la partie *j* ou à son indice de spécifité dans la même partie.

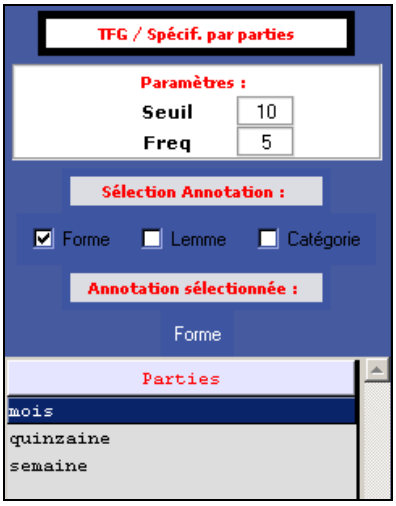


Figure 104 : Paramétrage TFG ; choix des paramètres de calcul, de l'annotation et de la partition

Pour lancer ce calcul, il faut :

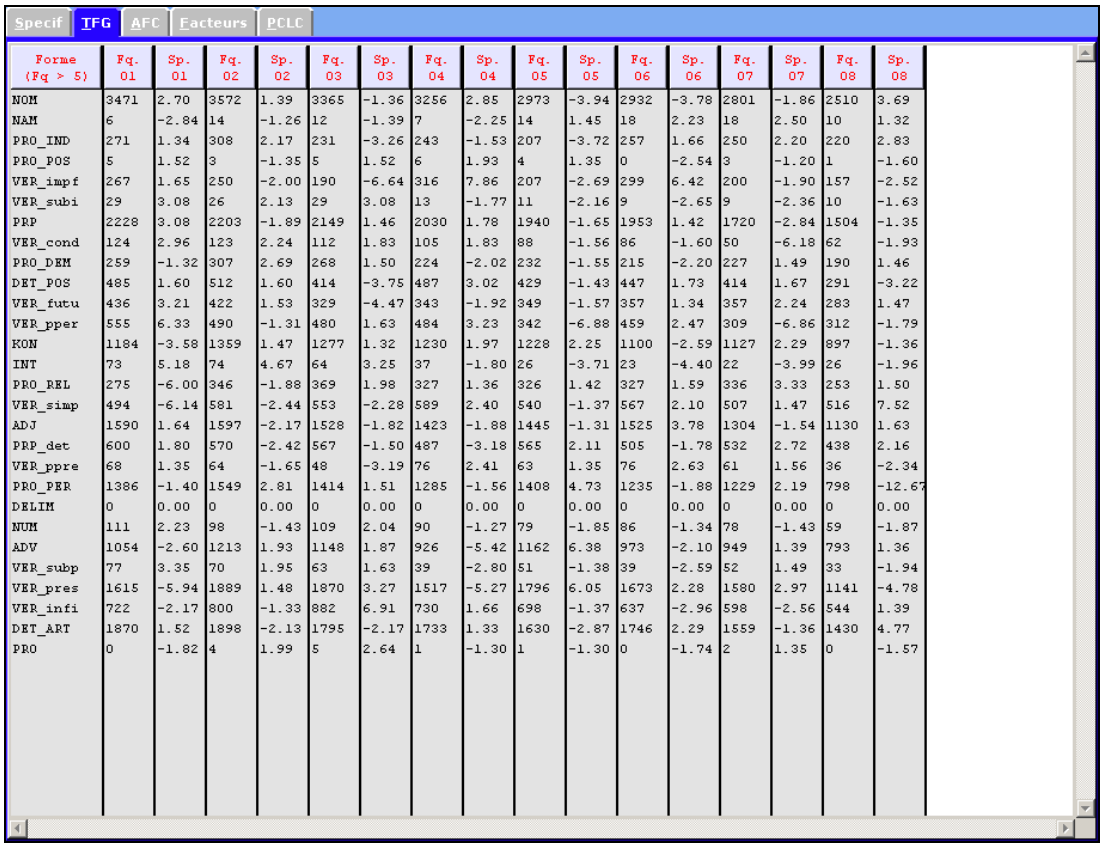
- sélectionner le niveau d'annotation visé ;
- sélectionner une des partitions disponibles dans la liste fournie sur la partie gauche de l'onglet **SPECIF** ;
- fixer les valeurs des paramètres (fréquence, seuil) ;

L'activation du bouton déclenche ensuite le calcul dont les résultats s'affichent dans l'onglet **TFG**

Specif | TFG | AFC | Facteurs | PCLC

Forme (Fq > 5)	Fq. 01	Sp. 01	Fq. 02	Sp. 02	Fq. 03	Sp. 03	Fq. 04	Sp. 04	Fq. 05	Sp. 05	Fq. 06	Sp. 06	Fq. 07	Sp. 07	Fq. 08	Sp. 08
accaparent	1	1.29	2	1.82	0	-1.32	1	1.31	0	-1.29	0	-1.29	1	1.34	0	-1.22
accapareur	0	-1.38	1	-1.10	1	1.23	0	-1.35	1	1.26	0	-1.34	0	-1.32	3	2.85
accapareurs	18	2.71	20	3.10	16	2.16	9	-1.36	2	-3.76	2	-3.70	10	1.36	3	-2.35
accaparé	0	-1.51	5	3.64	2	1.53	1	-1.14	0	-1.47	0	-1.46	0	-1.42	0	-1.35
accepter	1	1.29	3	2.63	0	-1.32	0	-1.29	0	-1.29	1	1.32	0	-1.26	0	-1.22
acceptée	4	3.40	1	-1.10	0	-1.38	1	1.25	0	-1.35	0	-1.34	0	-1.32	0	-1.26
accord	3	1.39	1	-1.56	2	-1.23	2	-1.20	1	-1.45	5	2.31	1	-1.39	2	1.31
accusant	1	1.29	1	1.27	0	-1.32	1	1.31	1	1.31	1	1.32	0	-1.26	0	-1.22
accusation	2	1.53	0	-1.54	3	2.08	0	-1.47	1	-1.13	1	-1.13	0	-1.42	1	1.26
accuse	0	-1.76	0	-1.80	3	1.67	2	1.34	1	-1.26	4	2.29	0	-1.63	2	1.49
accusent	2	-1.15	0	-1.94	0	-1.89	0	-1.82	2	1.27	6	3.36	1	-1.29	3	1.84
accuser	0	-1.63	0	-1.67	0	-1.63	0	-1.59	2	1.44	7	5.43	0	-1.53	1	-1.12
accusé	2	1.35	2	1.32	0	-1.70	1	-1.23	1	-1.23	3	1.84	2	1.44	0	-1.48
accusés	1	-1.19	0	-1.60	0	-1.57	0	-1.53	0	-1.52	3	2.05	0	-1.48	5	4.13
acheter	1	-1.22	0	-1.67	2	1.40	1	-1.20	1	-1.20	0	-1.57	4	2.69	1	-1.12
achever	4	1.96	0	-1.94	3	1.53	3	1.59	0	-1.81	2	1.28	1	-1.29	1	-1.22
achevé	1	1.29	0	-1.34	1	1.29	0	-1.29	1	1.31	1	1.32	1	1.34	0	-1.22
acte	1	1.29	0	-1.34	2	1.86	0	-1.29	0	-1.29	0	-1.29	1	1.34	1	1.40
action	3	2.24	0	-1.47	0	-1.44	0	-1.41	3	2.33	0	-1.40	0	-1.37	1	1.30
actions	2	1.35	0	-1.74	1	-1.26	0	-1.65	3	1.82	3	1.84	1	-1.20	1	-1.15
adam	1	-1.15	0	-1.54	0	-1.51	0	-1.47	2	1.58	0	-1.46	4	3.08	1	1.26
adieu	0	-1.32	0	-1.34	0	-1.32	1	1.31	0	-1.29	1	1.32	0	-1.26	3	3.12
administrati	0	-1.32	0	-1.34	3	2.70	1	1.31	0	-1.29	0	-1.29	1	1.34	0	-1.22
adresse	2	1.61	2	1.58	2	1.61	0	-1.41	0	-1.41	0	-1.40	0	-1.37	1	1.30
affaire	0	-1.44	0	-1.47	2	1.61	2	1.66	0	-1.41	2	1.68	0	-1.37	1	1.30
affaires	3	1.53	2	-1.17	0	-1.89	4	2.04	1	-1.34	0	-1.80	3	1.68	1	-1.22
affamer	1	-1.42	3	1.44	5	2.37	0	-1.88	0	-1.87	0	-1.86	4	2.08	2	1.37
affameurs	0	-1.44	1	-1.13	2	1.61	0	-1.41	0	-1.41	0	-1.40	3	2.43	1	1.30
affreuse	2	1.72	1	-1.10	2	1.72	0	-1.35	0	-1.35	0	-1.34	1	1.29	0	-1.26
affreux	2	1.30	1	-1.33	1	-1.30	0	-1.71	2	1.34	3	1.75	1	-1.23	2	1.49
afin	12	3.00	4	-1.61	7	1.44	3	-1.72	5	-1.26	4	-1.43	3	-1.59	5	1.40
agents	2	1.61	1	-1.13	0	-1.44	0	-1.41	0	-1.41	2	1.68	2	1.73	0	-1.31
ah	1	-3.66	10	1.50	5	-1.73	9	1.51	12	2.27	8	1.37	5	-1.48	8	1.72
ai	26	-1.36	21	-2.19	18	-2.54	36	2.67	30	1.73	48	6.26	14	-2.65	9	-3.28
aide	0	-1.38	1	-1.10	1	1.23	2	1.77	0	-1.35	1	1.26	1	1.29	0	-1.26
aient	3	1.60	1	-1.37	2	1.27	1	-1.30	1	-1.30	1	-1.29	2	1.35	2	1.45

Figure 105 : Tableau lexical (annotation n°1 : forme)



Forme (Fq > 5)	Fq. 01	Sp. 01	Fq. 02	Sp. 02	Fq. 03	Sp. 03	Fq. 04	Sp. 04	Fq. 05	Sp. 05	Fq. 06	Sp. 06	Fq. 07	Sp. 07	Fq. 08	Sp. 08
NOM	3471	2.70	3572	1.39	3365	-1.36	3256	2.85	2973	-3.94	2932	-3.78	2801	-1.86	2510	3.69
NAM	6	-2.84	14	-1.26	12	-1.39	7	-2.25	14	1.45	18	2.23	18	2.50	10	1.32
PRO_IND	271	1.34	308	2.17	231	-3.26	243	-1.53	207	-3.72	257	1.66	250	2.20	220	2.83
PRO_POS	5	1.52	3	-1.35	5	1.52	6	1.93	4	1.35	0	-2.54	3	-1.20	1	-1.60
VER_impf	267	1.65	250	-2.00	190	-6.64	316	7.86	207	-2.69	299	6.42	200	-1.90	157	-2.52
VER_subi	29	3.08	26	2.13	29	3.08	13	-1.77	11	-2.16	9	-2.65	9	-2.36	10	-1.63
PRP	2228	3.08	2203	-1.89	2149	1.46	2030	1.78	1940	-1.65	1953	1.42	1720	-2.84	1504	-1.35
VER_cond	124	2.96	123	2.24	112	1.83	105	1.83	88	-1.56	86	-1.60	50	-6.18	62	-1.93
PRO_DEM	259	-1.32	307	2.69	268	1.50	224	-2.02	232	-1.55	215	-2.20	227	1.49	190	1.46
DET_POS	485	1.60	512	1.60	414	-3.75	487	3.02	429	-1.43	447	1.73	414	1.67	291	-3.22
VER_futu	436	3.21	422	1.53	329	-4.47	343	-1.92	349	-1.57	357	1.34	357	2.24	283	1.47
VER_pper	555	6.33	490	-1.31	480	1.63	484	3.23	342	-6.88	459	2.47	309	-6.86	312	-1.79
KON	1184	-3.58	1359	1.47	1277	1.32	1230	1.97	1228	2.25	1100	-2.59	1127	2.29	897	-1.36
INT	73	5.18	74	4.67	64	3.25	37	-1.80	26	-3.71	23	-4.40	22	-3.99	26	-1.96
PRO_REL	275	-6.00	346	-1.88	369	1.98	327	1.36	326	1.42	327	1.59	336	3.33	253	1.50
VER_simp	494	-6.14	581	-2.44	553	-2.28	589	2.40	540	-1.37	567	2.10	507	1.47	516	7.52
ADJ	1590	1.64	1597	-2.17	1528	-1.82	1423	-1.88	1445	-1.31	1525	3.78	1304	-1.54	1130	1.63
PRP_det	600	1.80	570	-2.42	567	-1.50	487	-3.18	565	2.11	505	-1.78	532	2.72	438	2.16
VER_ppre	68	1.35	64	-1.65	48	-3.19	76	2.41	63	1.35	76	2.63	61	1.56	36	-2.34
PRO_PER	1386	-1.40	1549	2.81	1414	1.51	1285	-1.56	1408	4.73	1235	-1.88	1229	2.19	798	-12.67
DELIM	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
NUM	111	2.23	98	-1.43	109	2.04	90	-1.27	79	-1.85	86	-1.34	78	-1.43	59	-1.87
ADV	1054	-2.60	1213	1.93	1148	1.87	926	-5.42	1162	6.38	973	-2.10	949	1.39	793	1.36
VER_subp	77	3.35	70	1.95	63	1.63	39	-2.80	51	-1.38	39	-2.59	52	1.49	33	-1.94
VER_pres	1615	-5.94	1889	1.48	1870	3.27	1517	-5.27	1796	6.05	1673	2.28	1580	2.97	1141	-4.78
VER_infi	722	-2.17	800	-1.33	882	6.91	730	1.66	698	-1.37	637	-2.96	598	-2.56	544	1.39
DET_ART	1870	1.52	1898	-2.13	1795	-2.17	1733	1.33	1630	-2.87	1746	2.29	1559	-1.36	1430	4.77
PRO	0	-1.82	4	1.99	5	2.64	1	-1.30	1	-1.30	0	-1.74	2	1.35	0	-1.57

Figure 106 : Tableau lexical (annotation n°3 : catégorie)

Pour chaque ligne, associée à un item donné, on trouve successivement, la fréquence et l'indice de spécificité de cet item dans toutes les parties de la partition sélectionnée.

20.2 AFC : Analyse Factorielle des Correspondances

Le bouton

permet de réaliser une *analyse factorielle des correspondances* sur l'ensemble des parties du corpus. Le paramétrage permet de fixer entre autres le nombre des unités textuelles prises en compte dans l'analyse. Par défaut, l'analyse prend en compte les unités dont la fréquence est supérieure à 10.

Pour lancer ce calcul, il faut :

- sélectionner le niveau d'annotation visé ;
- sélectionner une des partitions disponibles dans la liste fournie sur la partie gauche de l'onglet **SPECIF** ;
- fixer les valeurs des paramètres (fréquence, seuil) ;

L'activation du bouton déclenche ensuite le calcul dont les résultats s'affichent dans l'onglet **AFC**. Les parties du corpus apparaissent sur le plan des deux premiers axes factoriels extraits par l'analyse.

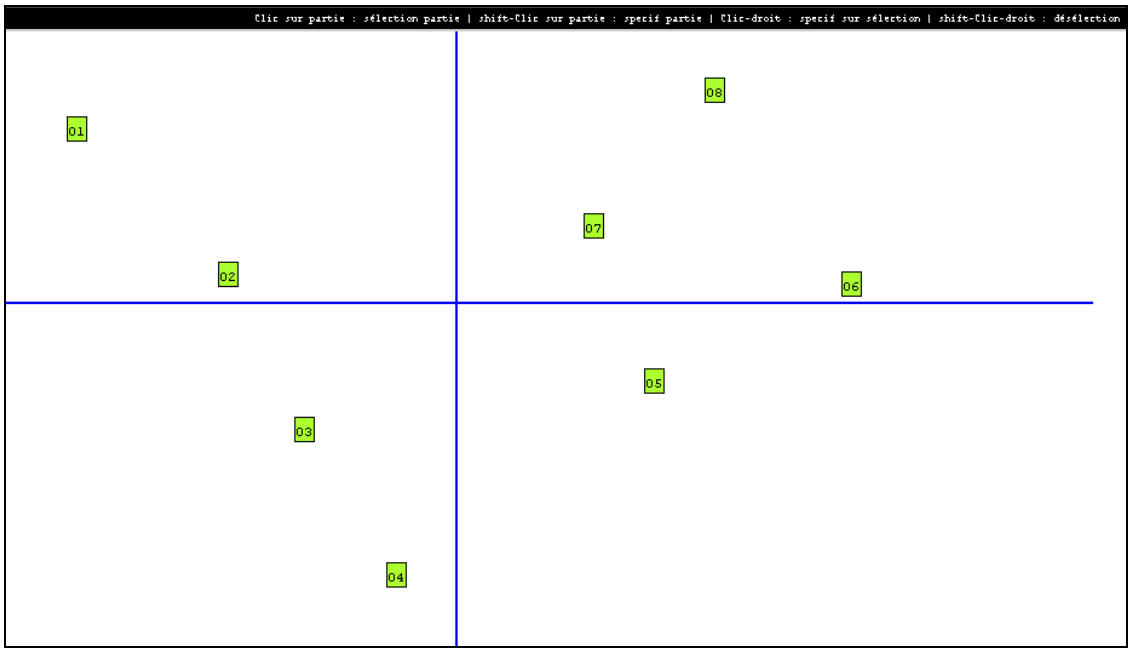


Figure 107 : AFC

En positionnant la souris sur une des parties, ses coordonnées dans le graphique sont affichées :

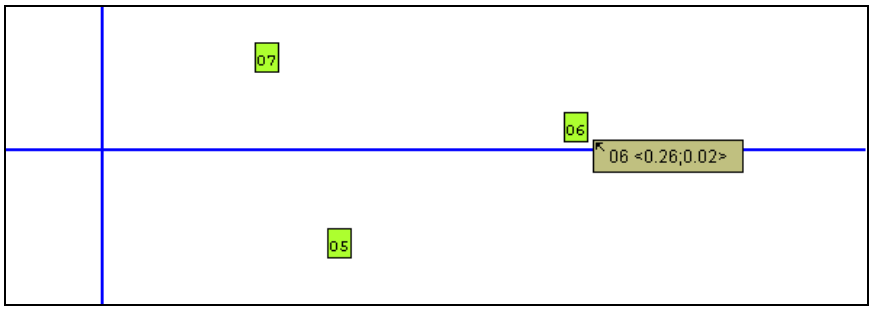


Figure 108 : AFC (position)

On peut ensuite sélectionner (clic gauche), directement sur la carte produite, une partie ou un groupe de parties. Les contours des parties sélectionnées apparaissent alors en surbrillance rouge. Cette sélection permet de calculer des spécificités sur le groupe de parties sélectionnées *via* un clic-droit sur le graphique.

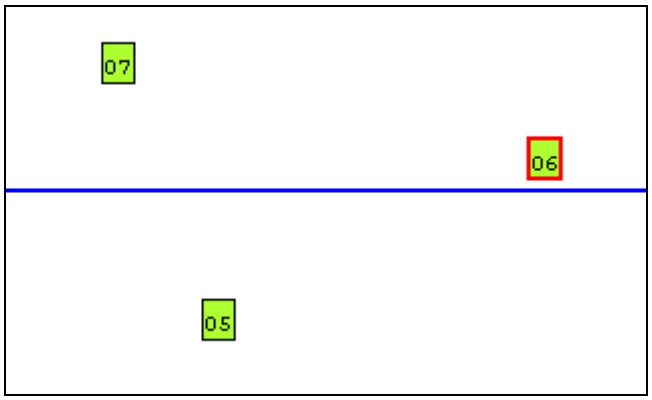


Figure 109 : AFC (sélection)

On peut aussi directement activer le calcul des spécificités d'une partie donnée *via* la combinaison de touche *shift+clic* sur la partie visée.

Le calcul de l'AFC peut aussi être lancé avec une option complémentaire permettant d'afficher dans le graphique final les dix items les plus spécifiques sur la partie visée. Pour activer cette option, il faut cocher la case AFC+TFG :

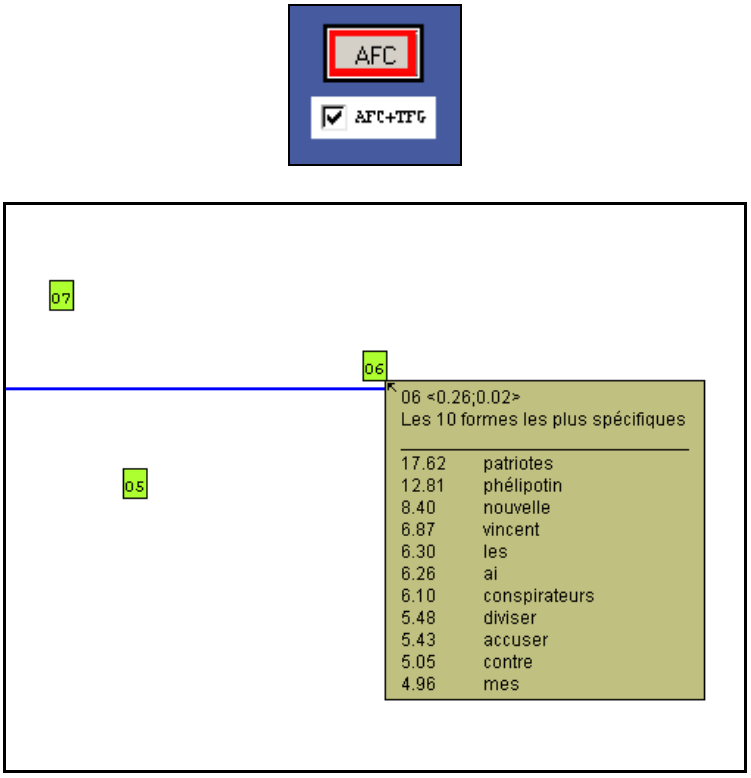


Figure 110 : AFC (affichage des mots spécifiques de la partie)

Le calcul de l'AFC sur une partition donnée déclenche aussi l'affichage :

- dans l'onglet **FACTEURS** : des facteurs construits lors du calcul de l'AFC permettant de résumer (de décrire brièvement) les variables actives initiales ;

Specif | TFG | AFC | Facteurs | PCLC

Partie	Masse	istanc	f1	f2	f3	f4	f5	f6	f1	f2	f3	f4	f5	f6	f1	f2	f3	f4	f5	f6
01	0.14	0.14	-0.25	0.18	0.15	-0.10	-0.04	0.00	31.50	19.60	14.20	7.16	1.15	0.02	0.45	0.23	0.15	0.07	0.01	0.00
02	0.14	0.11	-0.15	0.03	-0.03	0.11	0.20	-0.03	11.20	0.71	0.70	7.97	34.30	1.00	0.21	0.01	0.01	0.10	0.38	0.01
03	0.14	0.12	-0.10	-0.13	-0.02	0.16	-0.24	0.04	5.09	9.69	0.25	18.50	47.90	1.40	0.09	0.14	0.00	0.23	0.50	0.01
04	0.13	0.15	-0.04	-0.28	-0.06	-0.25	0.04	0.00	0.81	42.70	2.40	40.00	1.52	0.00	0.01	0.53	0.03	0.42	0.01	0.00
05	0.13	0.12	0.13	-0.08	0.06	0.15	0.06	-0.19	7.67	3.59	2.08	13.80	2.89	32.10	0.14	0.06	0.03	0.18	0.03	0.31
06	0.12	0.16	0.26	0.02	0.25	-0.04	0.01	0.11	30.80	0.25	37.20	0.82	0.02	11.10	0.44	0.00	0.41	0.01	0.00	0.08
07	0.11	0.14	0.09	0.08	-0.22	0.06	0.07	0.22	3.12	3.31	25.70	2.40	2.93	36.30	0.06	0.05	0.35	0.03	0.03	0.34
08	0.09	0.19	0.17	0.22	-0.20	-0.14	-0.13	-0.17	9.79	20.20	17.40	9.41	9.23	18.10	0.15	0.26	0.21	0.10	0.09	0.15

Figure 111 : Les facteurs de l'AFC

- dans l'onglet **PCLC** : des principales caractéristiques lexicométriques des différentes parties de la partition sélectionnée.

Specif | TFG | AFC | Facteurs | PCLC

Partition	Valeur Partie	début	fin	fq item (forme)	FQ item (occur)
mois	01	1	41214	3439	19256
mois	02	41215	85366	3778	20343
mois	03	85367	126728	3563	19278
mois	04	126729	165735	3542	18004
mois	05	165736	204892	3373	17814
mois	06	204893	243793	3392	17545
mois	07	243794	279786	3297	16295
mois	08	279787	310144	2970	13644

Figure 112 : PCLC d'une partition

Un premier calcul de l'AFC ayant été lancé, il est possible de relancer le calcul en sélectionnant des parties de la partition visee à ne pas prendre en compte pour le calcul : les parties selectionnées dans la liste des PCLC ne seront pas integrées au calcul.

20.3 Intégration de modules R (phase 1)

Cette partie décrit la mise à jour mise en œuvre à partir de la version 11.00 du ***Trameur*** permettant de lancer des modules R (AFC, ACP, CAH) via la librairie *factoMineR* (http://factominer.free.fr/index_fr.html)

Pour illustrer ces fonctionnalités, on utilise ici la base textométrique intitulée « Base des états finaux du corpus BROUILLONS (ANR ECRITURES[19]) » :

Cette base rassemble 19 « états finaux » issus de 19 dossiers[20] du corpus BROUILLONS (ANR ECRITURES). Chaque item de la *Trame* est associé à 3 niveaux d'annotation : forme, lemme, catégorie. On présente ci-dessous un extrait de la *Trame* de cette base via une concordance :

```
------------------------------PARTIE{dossier=ravet}------------------------------
        un lycée professionnel standard mais nous craignons que adèle ne
             compte tenu de ce que nous connaiss
      relation avec les garçons que nous trouvi
             compte tenu de ce que nous observ
               un autre garçon accueilli) nous
                 . § a ce jour nous n'avons aucu
                  vacances et elle a pu nous dire que sa grand-
          madame richard, assistante familiale, nous a informé que adèle
                    de sa famille. § nous avons reçu adèle pour
                      de la dap). § nous lui avons rappelé qu'
                  famille non préparé. § nous  allons proposer à la
                  fragilités de adèle, il nous semble qu'un accueil
         l'institut camille lamartine, nous avons organisé des rencontres
               adèle a été demandeuse .nous allons en organiser une
            foyer des oiseaux. § nous sommes actuellement en contact
                    elle.  § toutefois, nous pensons qu'il y
```

Position:<482>
Forme:<nous>|Freq:317
Lemme:<Nous_Sujet_Sentiment>|Freq:13
Cat:<Nous_Sujet_Sentiment>|Freq:13

Les items associés à la forme (annotation n°1) « nous » ont été annotées via le ***Trameur*** : leur catégorie (annotation n°3) a été modifiée pour prendre les valeurs suivantes :

[19] http://www.univ-paris3.fr/anr-ecritures-96530.kjsp?RH=1295620557102

[20] Chaque dossier est initialement constitué d'un ensemble chronologique de textes allant du brouillon initial à la version finale, on ne retient ici que ce dernier état.

```
Nous_Faire
Nous_OBJ
Nous_Sujet
Nous_Sujet_Constat
Nous_Sujet_Dire
Nous_Sujet_Dire_Faire
Nous_Sujet_Dire_Preco
Nous_Sujet_Faire
Nous_Sujet_Pensee
Nous_Sujet_Pensee_Faire
Nous_Sujet_Preco
Nous_Sujet_Savoir
Nous_Sujet_Sentiment
```

Mise en œuvre des modules R dans le Trameur

Les modules R sont disponibles dans l'onglet **STAT** qui a désormais l'allure suivante :

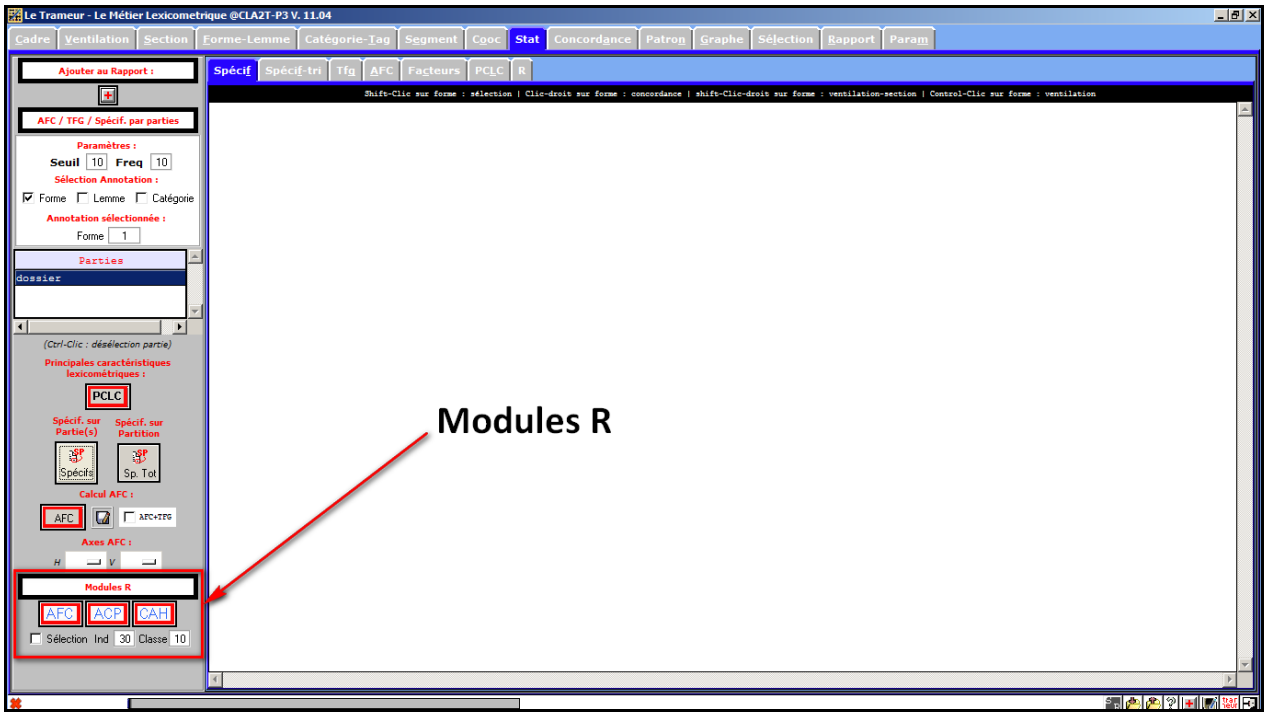


Figure 113 : Modules R (1)

Le fichier de données utilisés par R est généré dynamiquement dans le dossier contenant la base de travail. Les 3 modules produisent des sorties PDF (rassemblant les graphiques produits par R) qui sont stockées dans le sous-dossier *export* du dossier contenant la base de travail.

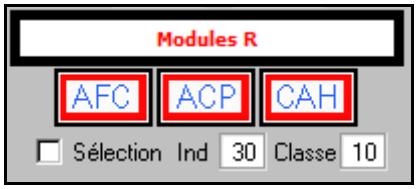


Figure 114 : Modules R (2)

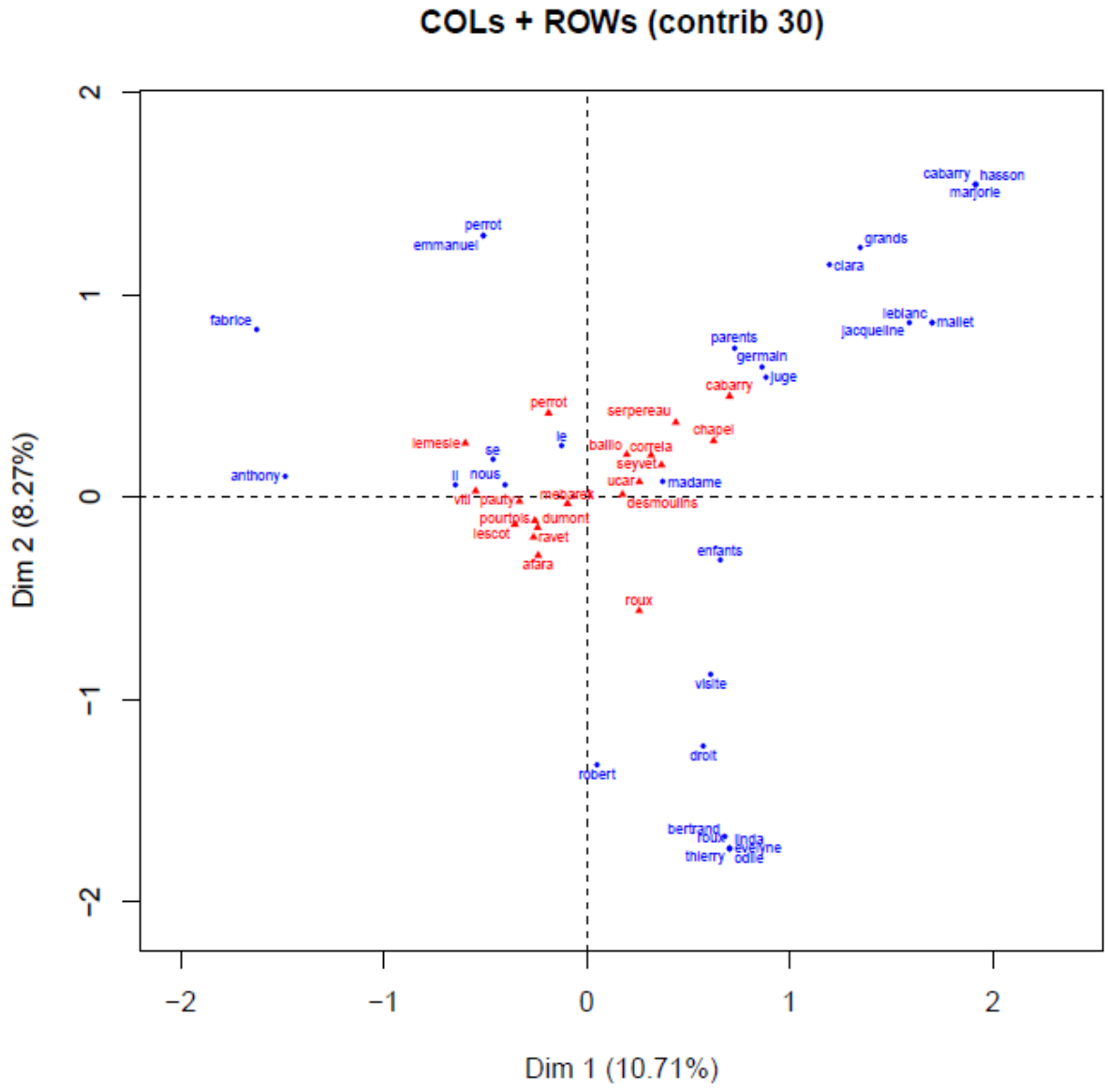
Par défaut, les 3 modules utilisent l'annotation sélectionnée dans le paramétrage de l'onglet **STAT** (en haut à gauche) et les items de fréquence supérieure à celle du paramètre *freq* (même zone de paramétrage, en haut à gauche).

- Le paramètre *Ind* permet de spécifier le nombre d'item (forme, lemme, catégorie…) qui seront projetés sur les graphiques de l'AFC ou de l'ACP.
- Le paramètre *Classe* permet de spécifier le nombre de classes à mettre au jour pour le module CAH.
- Le paramétre *Sélection* permet de modifier le tableau de données en entrée de R en ne gardant que les items préalablement sélectionnés via le *Gestionnaire de Sélection*.

Illustration sur la base « Etats finaux BROUILLONS »

Calcul sur base complète

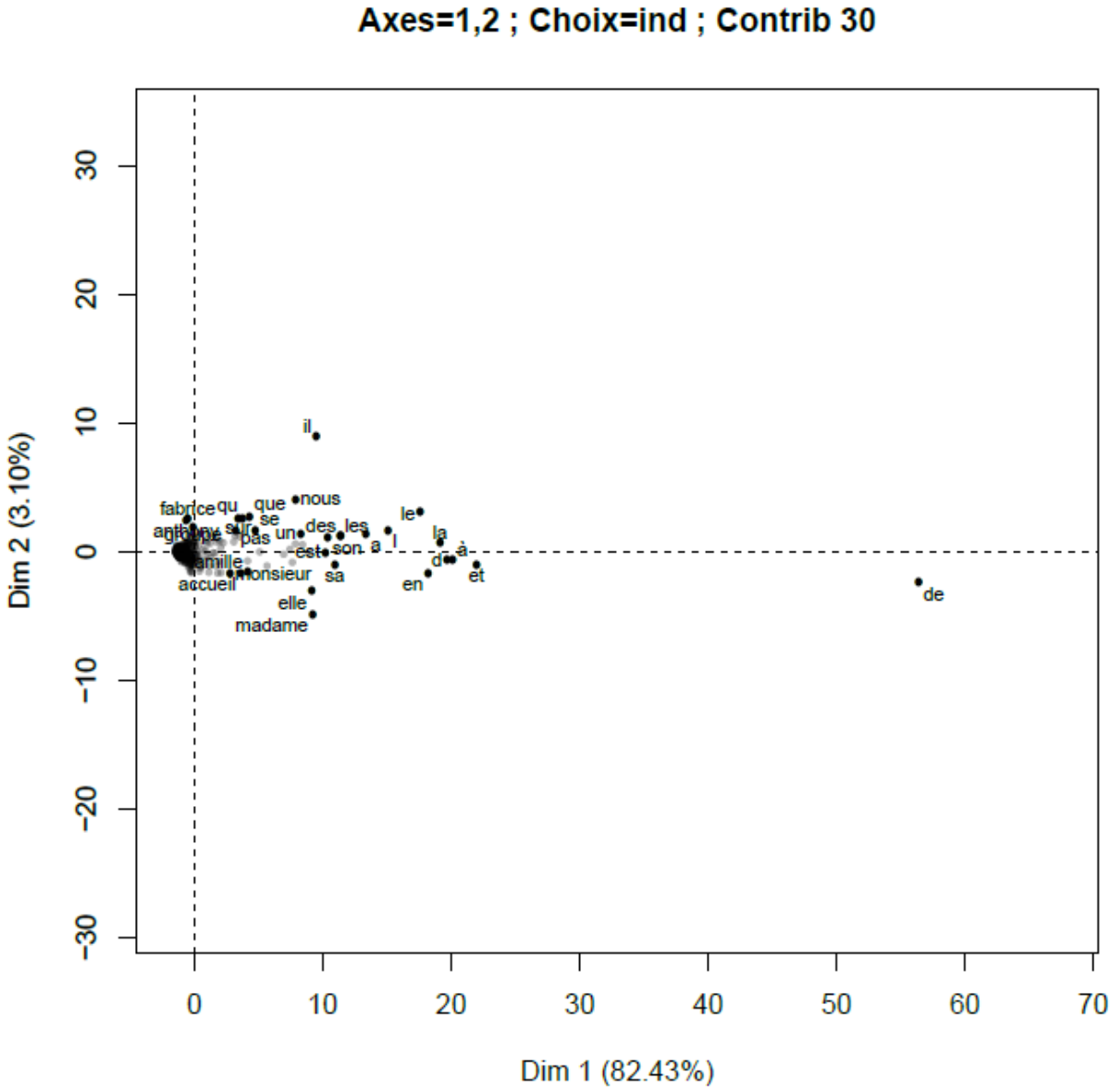
La base étant chargée, on active le calcul de l'AFC via R (paramétrage par défaut) via le bouton idoine, on donne à voir ci-dessous un extrait de la sortie PDF produite :




```
------ AFC OK ------
Fichier de résultats (PDF) : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export/AFC1375361223.pdf
Fichier de résultats (TXT) : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export/AFC-Result-1375361223.txt
Fichier de données : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/dataForR1375361223.txt
Répertoire de travail : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export
Répertoire de logs : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/logR
Répertoire de R : C:/SergeFleury/SF-WORK/07-DEVEL/ANR-textometrie/mkGraphTrameCadre/R/bin/x64/R.exe
```

Figure 115 : AFC via R

De même pour l'ACP :




```
------ ACP OK ------
Fichier de résultats (PDF) : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export/ACP1375361365.pdf
Fichier de résultats (TXT) : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export/ACP-Result-1375361365.txt
Fichier de données : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/dataForR1375361365.txt
Répertoire de travail : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export
Répertoire de logs : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/logR
Répertoire de R : C:/SergeFleury/SF-WORK/07-DEVEL/ANR-textometrie/mkGraphTrameCadre/R/bin/x64/R.exe
```

Figure 116 : ACP via R

De même pour le module CAH en modifiant ici le calcul pour travailler sur l'annotation n°3 (catégorie) :

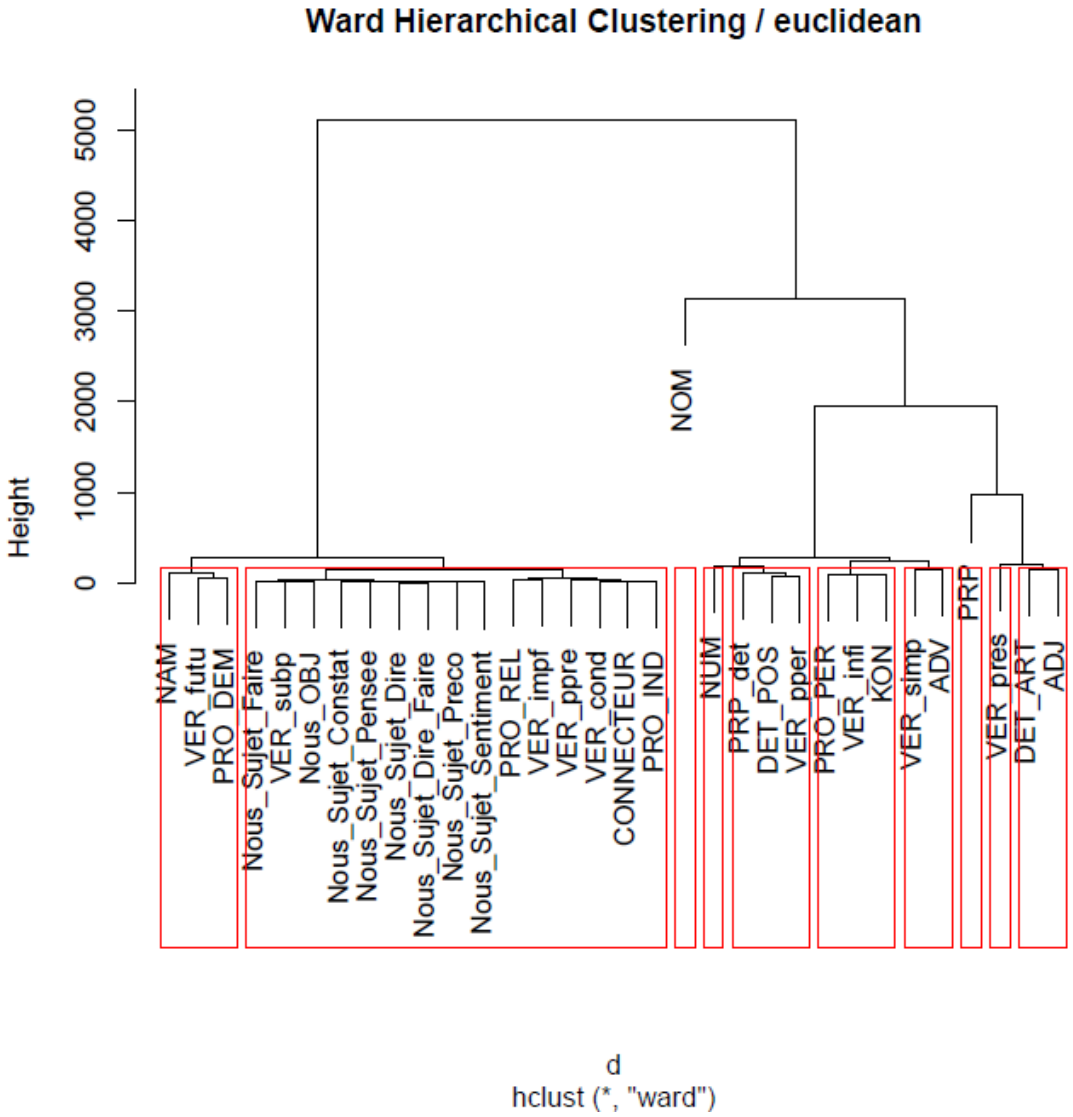


Figure 117 : CAH via R

Calcul via le *Gestionnaire de Sélection*

On souhaite désormais que les calculs soient faits uniquement sur les items de type VERBE et ceux correspondant aux différents types de « nous ».

Pour repérer ces items, on commence par les sélectionner via le *Gestionnaire de Sélection* (cf *infra*) : recherche en 2 temps, les « nous » puis les verbes (cf figures suivantes) :

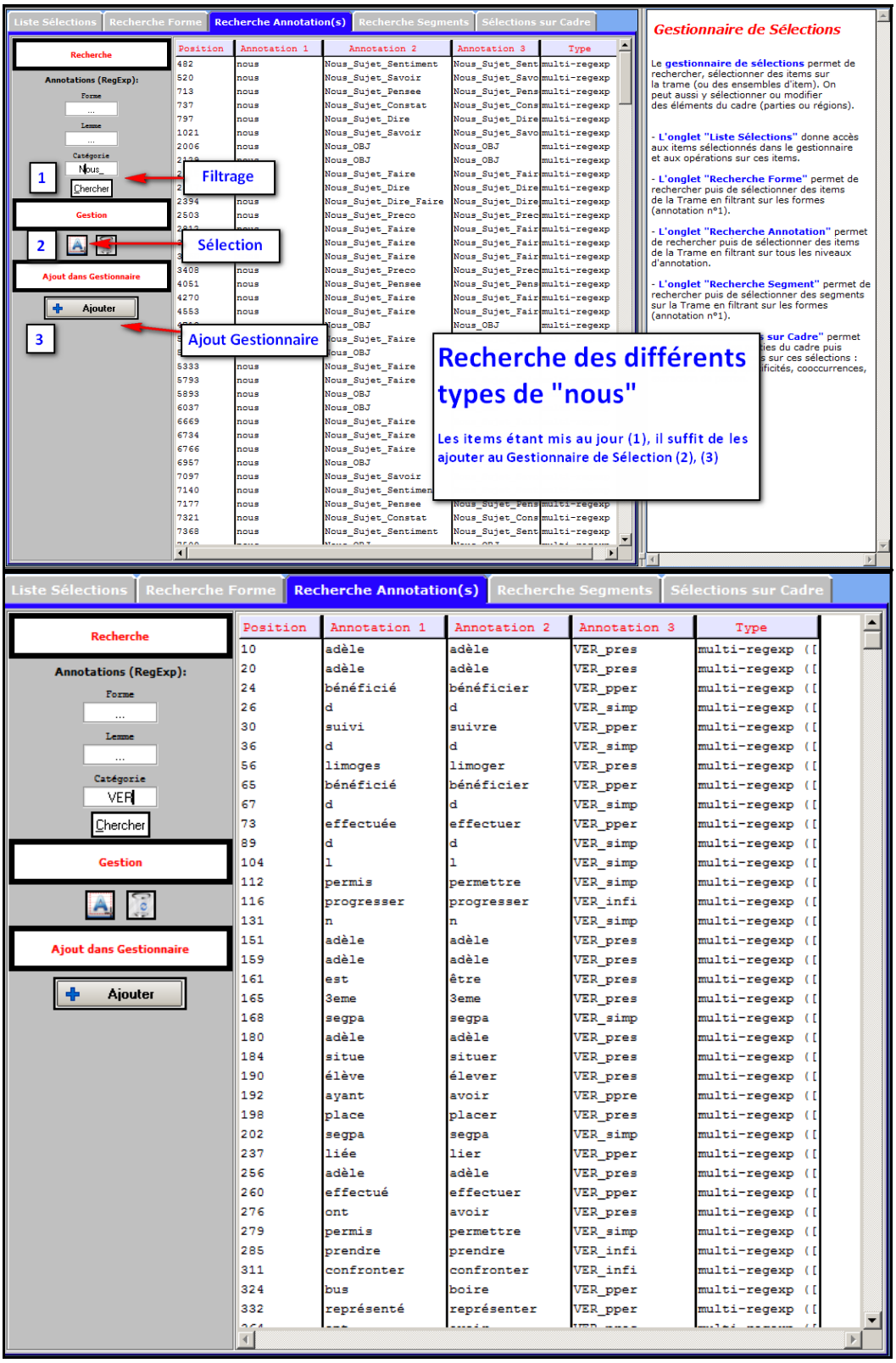

Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Sélections sur Cadre
Recherche
Annotations (RegExp):
Forme
Lemme
Catégorie
Nous_
Chercher
Gestion
Ajout dans Gestionnaire
Ajouter
1
2
3
Filtrage
Sélection
Ajout Gestionnaire
Position
Annotation 1
Annotation 2
Annotation 3
Type
Gestionnaire de Sélections
Le gestionnaire de sélections permet de rechercher, sélectionner des items sur la trame (ou des ensembles d'item). On peut aussi y sélectionner ou modifier des éléments du cadre (parties ou régions).
- L'onglet "Liste Sélections" donne accès aux items sélectionnés dans le gestionnaire et aux opérations sur ces items.
- L'onglet "Recherche Forme" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur les formes (annotation n°1).
- L'onglet "Recherche Annotation" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur tous les niveaux d'annotation.
- L'onglet "Recherche Segment" permet de rechercher puis de sélectionner des segments sur la Trame en filtrant sur les formes (annotation n°1).
Recherche des différents types de "nous"
Les items étant mis au jour (1), il suffit de les ajouter au Gestionnaire de Sélection (2), (3)
Catégorie
VER
Chercher
Ajout dans Gestionnaire
Ajouter

Figure 118 : Sélection des items pour les modules R

Au final, le *Gestionnaire de Sélection* contient tous les items recherchés, on prend soin de tous les sélectionner :

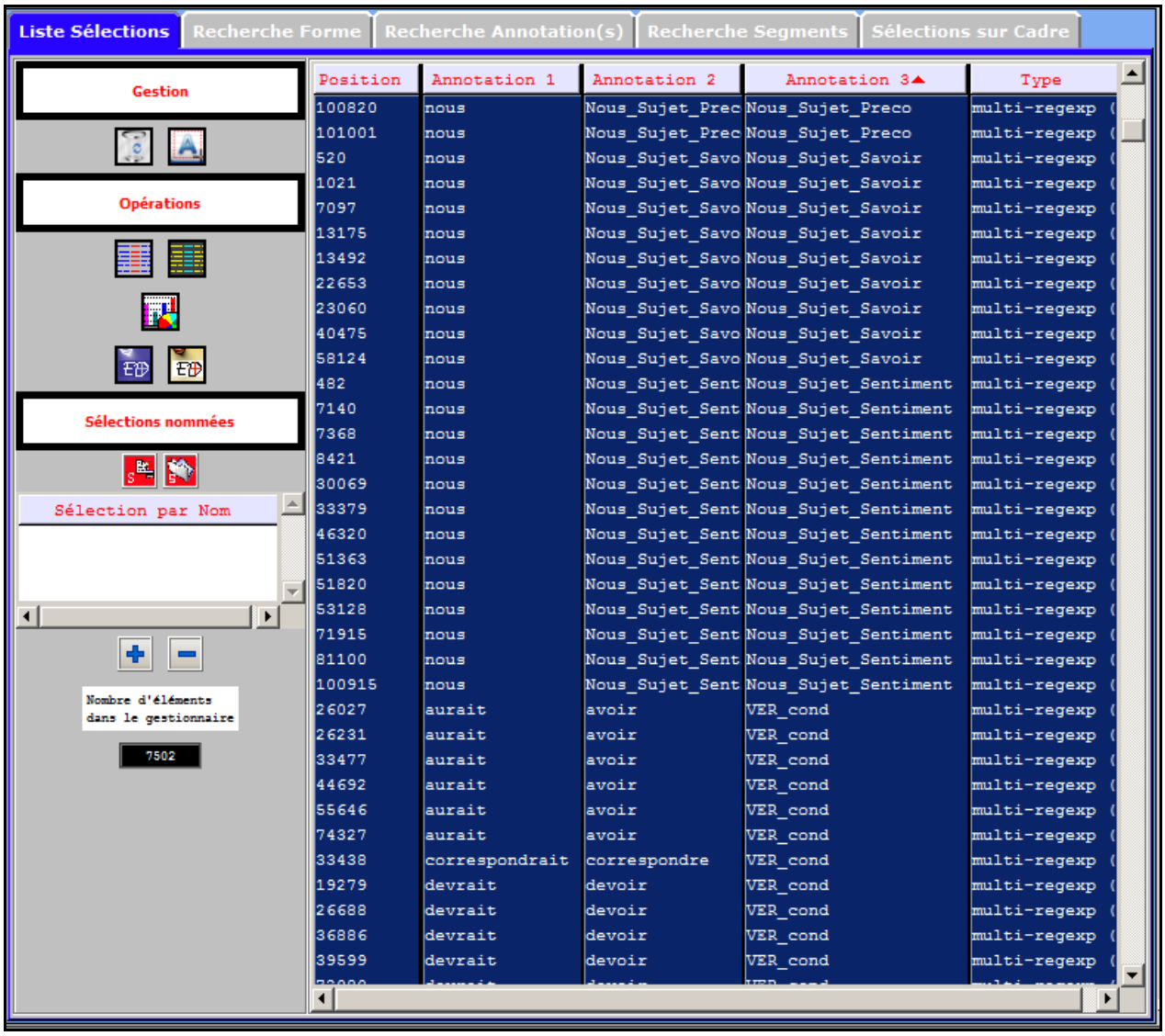

Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Sélections sur Cadre
Gestion
Opérations
Sélections nommées
Sélection par Nom
Nombre d'éléments dans le gestionnaire
7502
Position
Annotation 1
Annotation 2
Annotation 3
Type
100820 nous Nous_Sujet_Prec Nous_Sujet_Preco multi-regexp
101001 nous Nous_Sujet_Prec Nous_Sujet_Preco multi-regexp
520 nous Nous_Sujet_Savo Nous_Sujet_Savoir multi-regexp
1021 nous Nous_Sujet_Savo Nous_Sujet_Savoir multi-regexp
7097 nous Nous_Sujet_Savo Nous_Sujet_Savoir multi-regexp
13175 nous Nous_Sujet_Savo Nous_Sujet_Savoir multi-regexp
13492 nous Nous_Sujet_Savo Nous_Sujet_Savoir multi-regexp
22653 nous Nous_Sujet_Savo Nous_Sujet_Savoir multi-regexp
23060 nous Nous_Sujet_Savo Nous_Sujet_Savoir multi-regexp
40475 nous Nous_Sujet_Savo Nous_Sujet_Savoir multi-regexp
58124 nous Nous_Sujet_Savo Nous_Sujet_Savoir multi-regexp
482 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
7140 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
7368 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
8421 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
30069 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
33379 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
46320 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
51363 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
51820 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
53128 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
71915 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
81100 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
100915 nous Nous_Sujet_Sent Nous_Sujet_Sentiment multi-regexp
26027 aurait avoir VER_cond multi-regexp
26231 aurait avoir VER_cond multi-regexp
33477 aurait avoir VER_cond multi-regexp
44692 aurait avoir VER_cond multi-regexp
55646 aurait avoir VER_cond multi-regexp
74327 aurait avoir VER_cond multi-regexp
33438 correspondrait correspondre VER_cond multi-regexp
19279 devrait devoir VER_cond multi-regexp
26688 devrait devoir VER_cond multi-regexp
36886 devrait devoir VER_cond multi-regexp
39599 devrait devoir VER_cond multi-regexp

Figure 119 : Items sélectionnés dans la *Gestionnaire de Sélection*

Les items pertinents étant sélectionnés, on peut désormais lancer les modules R sur ces items uniquement :

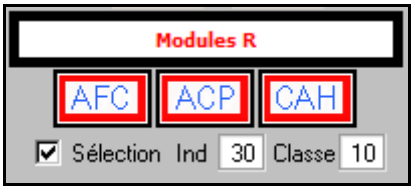


Figure 120 : Paramétrage R pour travailler avec le *Gestionnaire de Sélection*

Il peut être nécessaire/utile de modifier aussi les paramétres *Ind* et/ou *Classe*.

Calcul de l'AFC

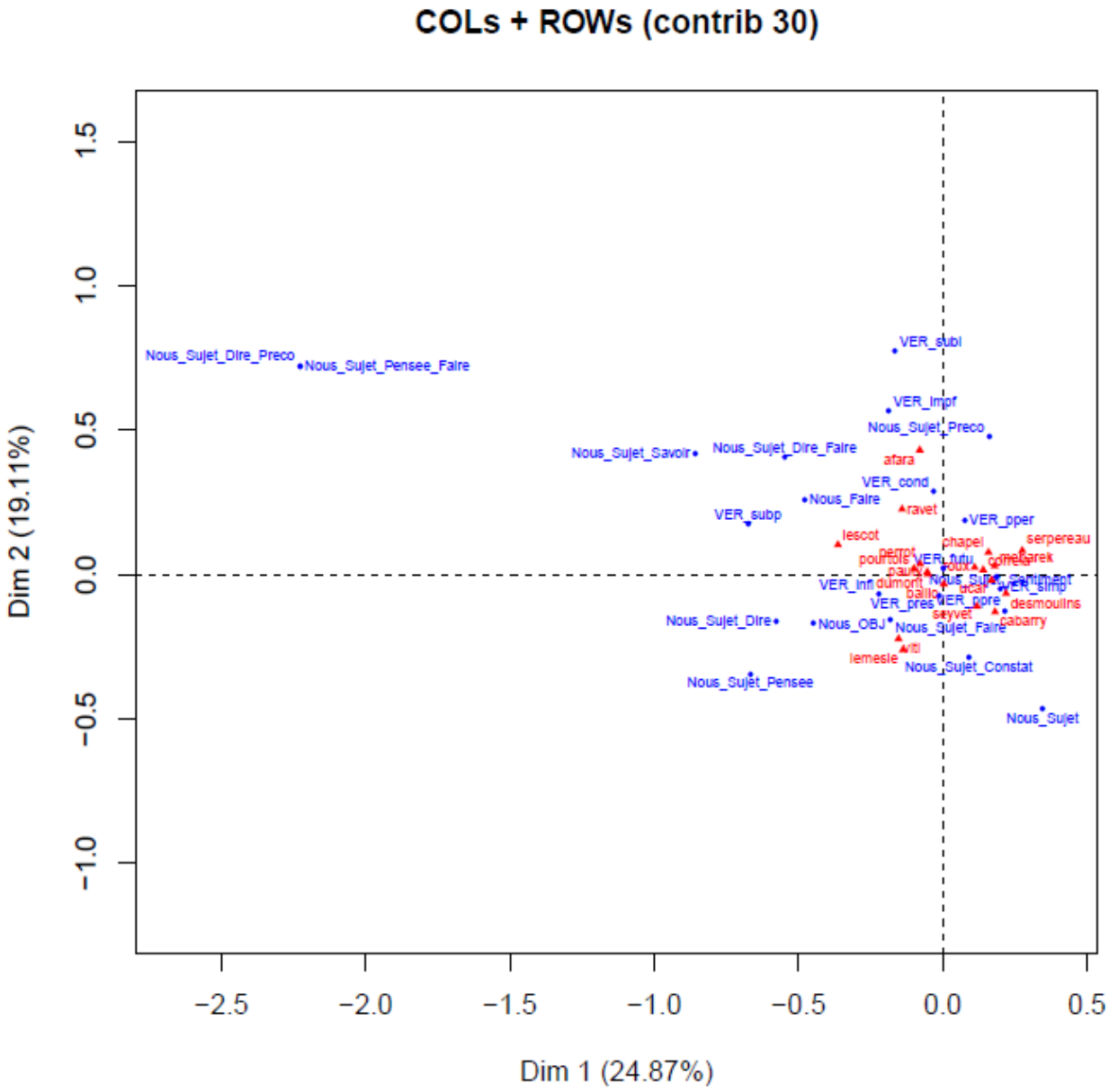


Figure 121 : AFC via R sur le contenu du *Gestionnaire de Sélection*

Calcul de la CAH

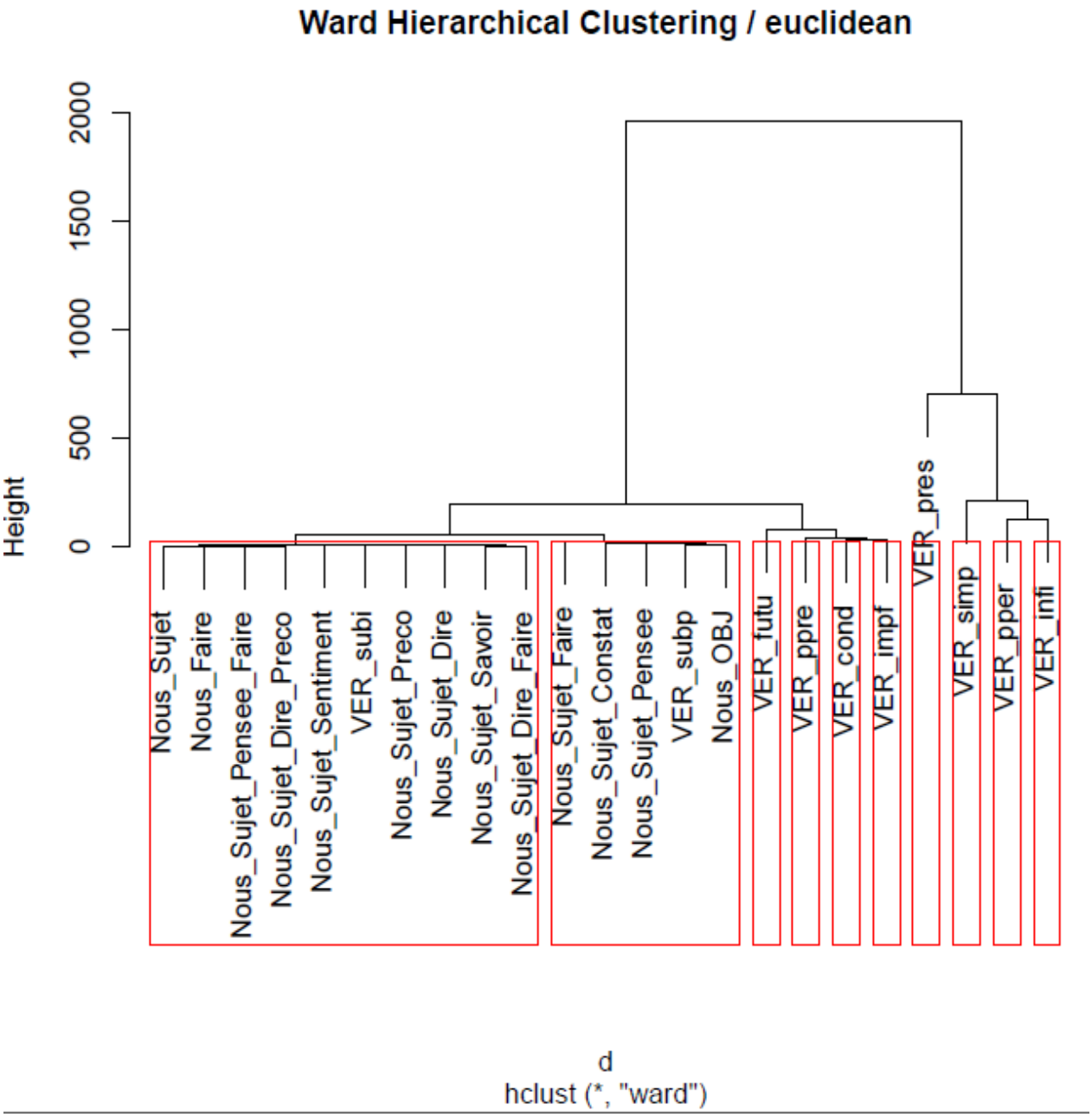


Figure 122 : CAH via R sur le contenu du *Gestionnaire de Sélection*

21 Sélection. *Gestionnaire de sélection*

> *La sélection est un objet textométrique qui permet de décrire de manière homogène la plupart des objets sur lesquels portent les opérations textométriques. Une sélection est un sous-ensemble d'items appartenant à la **Trame**.*
>
> *Pour construire des unités complexes et étudier leur répartition dans différentes zones d'un corpus de textes, un logiciel de textométrie doit disposer d'un gestionnaire de sélection permettant de constituer, de modifier, de gérer, les sélections à partir de toutes les vues possibles sur le texte et les sélections déjà construites. Le gestionnaire de sélections prend appui sur toutes les présentations possibles du texte ou des comptages d'unités réalisés à partir de ce dernier pour définir d'autres unités qui seront elles-mêmes soumises à des évaluations textométriques.*
>
> *(Söze-Duval 2008)*

Le ***Trameur*** dispose de fonctionnalités permettant de sélectionner des positions de la ***Trame***, de les mémoriser et de constituer des unités réutilisables dans des opérations de ventilation ou de concordance. L'onglet **SELECTION** enregistre au fur et à mesure toutes les sélections réalisées ; il donne aussi accès à l'outil *Gestionnaire de Sélection* qui permet de travailler à partir de ces sélections.

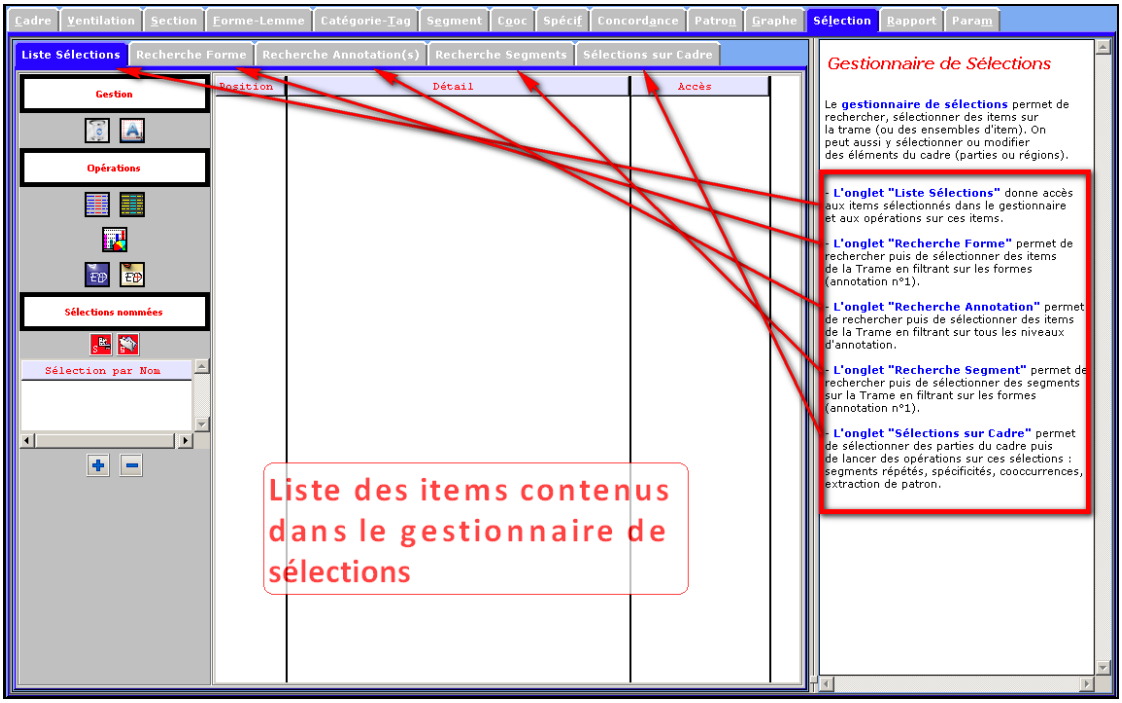

Cadre
Ventilation
Section
Forme-Lemme
Catégorie-Tag
Segment
Cooc
Spécif
Concordance
Patron
Graphe
Sélection
Rapport
Param
Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Sélections sur Cadre
Gestion
Opérations
Sélections nommées
Sélection par Nom
Position
Détail
Accès
Liste des items contenus dans le gestionnaire de sélections
Gestionnaire de Sélections
Le gestionnaire de sélections permet de rechercher, sélectionner des items sur la trame (ou des ensembles d'item). On peut aussi y sélectionner ou modifier des éléments du cadre (parties ou régions).
- L'onglet "Liste Sélections" donne accès aux items sélectionnés dans le gestionnaire et aux opérations sur ces items.
- L'onglet "Recherche Forme" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur les formes (annotation n°1).
- L'onglet "Recherche Annotation" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur tous les niveaux d'annotation.
- L'onglet "Recherche Segment" permet de rechercher puis de sélectionner des segments sur la Trame en filtrant sur les formes (annotation n°1).
- L'onglet "Sélections sur Cadre" permet de sélectionner des parties du cadre puis de lancer des opérations sur ces sélections : segments répétés, spécificités, cooccurrences, extraction de patron.

Figure 123 : *Le Trameur* onglet SELECTION

Le *Gestionnaire de Sélection* permet de « sélectionner» des items de la ***Trame*** (via leur position) ou de « sélectionner » des zones de la ***Trame*** (via les parties disponibles). Dans les deux cas, on peut ensuite lancer des opérations sur les sélections réalisées.

21.1 Sélection de positions sur la Trame

On distingue deux modes pour sélectionner des positions de la ***Trame*** :

- Le mode **sélection directe à partir de la *Trame*** : ce mode est disponible sur les objets *éditeur* (onglet **CADRE** et **VENTILATION**), l'objet *concordance* (onglet

CONCORDANCE) et les objets *dictionnaire* (onglets **FORME, LEMME, CATEGORIE**).

- Le mode **génération d'une sélection** : ce mode est accessible dans l'outil *Gestionnaire de Sélection*. Dans ce mode, il est possible de choisir dans un ensemble d'unités (formes, segments) vérifiant une propriété visée.

21.1.1 Sélection directe à partir de la *Trame*

21.1.1.1 Sélectionner dans l'éditeur du *Cadre*

On dispose dans l'onglet **CADRE** d'un raccourci clavier (*shift-clic*) permettant de sélectionner les positions visées. Après avoir édité la ***Trame*** d'une partie du ***Cadre*** (accès à la ***Trame*** par le ***Cadre***), les items de la ***Trame*** imprimés sont sélectionnables : l'activation du raccourci clavier sur un item déclenche (1) sa sélection (une trace de cette opération est visible dans l'onglet **SELECTION**) et (2) une coloration jaune autour de l'item visé :

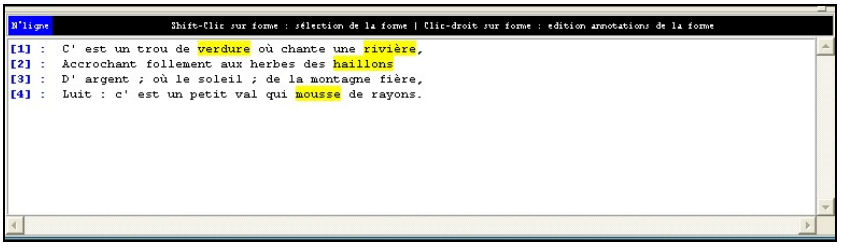

N°ligne
Shift-Clic sur forme : sélection de la forme | Clic-droit sur forme : edition annotations de la forme
[1] : C' est un trou de verdure où chante une rivière,
[2] : Accrochant follement aux herbes des haillons
[3] : D' argent ; où le soleil ; de la montagne fière,
[4] : Luit : c' est un petit val qui mousse de rayons.

Figure 124 : *Le Trameur* sélection sur la *Trame* (onglet CADRE)

Dans la figure précédente, les positions associées aux items surlignés en jaune sont désormais sélectionnées dans le *Gestionnaire de Sélection*.

21.1.1.2 Sélectionner dans l'éditeur des sections

Le même mécanisme est disponible dans l'éditeur de la ***Trame*** par les sections.

21.1.1.3 Sélectionner dans l'éditeur des graphiques

Un mécanisme similaire est disponible sur l'éditeur de l'onglet **VENTILATION** quand celui-ci est utilisé pour construire la courbe d'accroissement du vocabulaire. Cet éditeur affiche une trace du calcul de la courbe d'accroissement donnant à voir pour chaque position de la ***Trame*** les valeurs retenues à cette position. L'activation du raccourci clavier sur une position déclenche sa sélection.

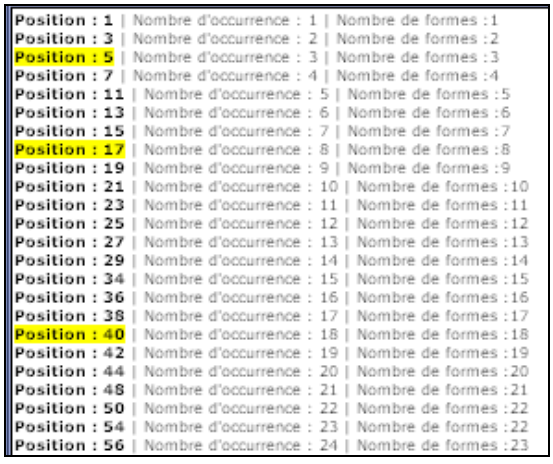

Position : 1 | Nombre d'occurrence : 1 | Nombre de formes :1
Position : 3 | Nombre d'occurrence : 2 | Nombre de formes :2
Position : 5 | Nombre d'occurrence : 3 | Nombre de formes :3
Position : 7 | Nombre d'occurrence : 4 | Nombre de formes :4
Position : 11 | Nombre d'occurrence : 5 | Nombre de formes :5
Position : 13 | Nombre d'occurrence : 6 | Nombre de formes :6
Position : 15 | Nombre d'occurrence : 7 | Nombre de formes :7
Position : 17 | Nombre d'occurrence : 8 | Nombre de formes :8
Position : 19 | Nombre d'occurrence : 9 | Nombre de formes :9
Position : 21 | Nombre d'occurrence : 10 | Nombre de formes :10
Position : 23 | Nombre d'occurrence : 11 | Nombre de formes :11
Position : 25 | Nombre d'occurrence : 12 | Nombre de formes :12
Position : 27 | Nombre d'occurrence : 13 | Nombre de formes :13
Position : 29 | Nombre d'occurrence : 14 | Nombre de formes :14
Position : 34 | Nombre d'occurrence : 15 | Nombre de formes :15
Position : 36 | Nombre d'occurrence : 16 | Nombre de formes :16
Position : 38 | Nombre d'occurrence : 17 | Nombre de formes :17
Position : 40 | Nombre d'occurrence : 18 | Nombre de formes :18
Position : 42 | Nombre d'occurrence : 19 | Nombre de formes :19
Position : 44 | Nombre d'occurrence : 20 | Nombre de formes :20
Position : 48 | Nombre d'occurrence : 21 | Nombre de formes :21
Position : 50 | Nombre d'occurrence : 22 | Nombre de formes :22
Position : 54 | Nombre d'occurrence : 23 | Nombre de formes :22
Position : 56 | Nombre d'occurrence : 24 | Nombre de formes :23

Figure 125 : *Le Trameur* sélection sur la *Trame* (onglet VENTILATION)

21.1.1.4 Sélectionner dans les concordances

Le même mécanisme de sélection est disponible dans l'onglet **CONCORDANCE** sur les pôles des concordances construites. Le même raccourci clavier (*shift-clic*) permet de sélectionner la position de la forme pôle visée :

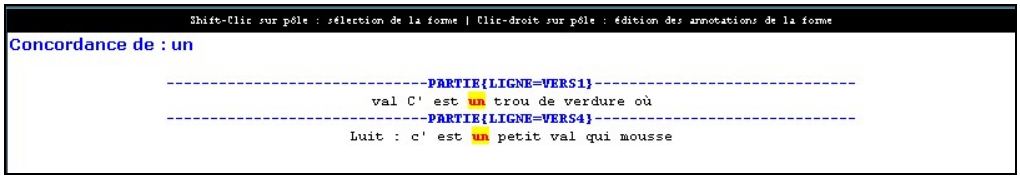

Shift-Clic sur pôle : sélection de la forme | Clic-droit sur pôle : édition des annotations de la forme
Concordance de : un
PARTIE{LIGNE=VERS1}
val C' est un trou de verdure où
PARTIE{LIGNE=VERS4}
Luit : c' est un petit val qui mousse

Figure 126 : *Le Trameur* sélection sur une concordance (onglet CONCORDANCE)

21.1.2 Sélectionner dans les dictionnaires

On peut aussi sélectionner les ensembles de positions associées aux entrées des différents dictionnaires (onglets **FORME**, **LEMME**, **CATEGORIE**).

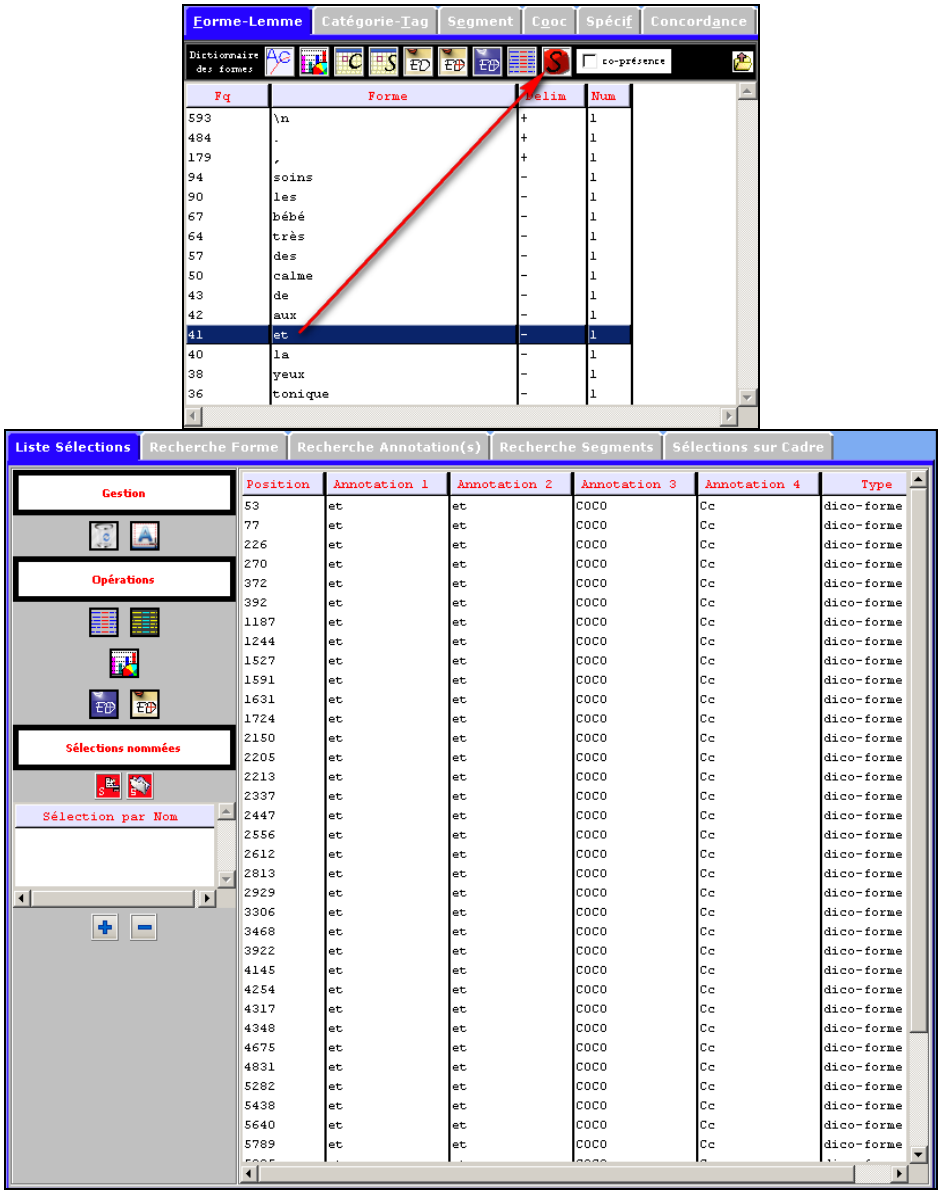

Forme-Lemme
Catégorie-Tag
Segment
Cooc
Spécif
Concordance
Dictionnaire des formes
co-présence
Fq
Forme
Delim
Num
593 \n + 1
484 . + 1
179 , + 1
94 soins - 1
90 les - 1
67 bébé - 1
64 très - 1
57 des - 1
50 calme - 1
43 de - 1
42 aux - 1
41 et - 1
40 la - 1
38 yeux - 1
36 tonique - 1
Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Sélections sur Cadre
Gestion
Opérations
Sélections nommées
Sélection par Nom
Position
Annotation 1
Annotation 2
Annotation 3
Annotation 4
Type
53 et et COCO Cc dico-forme
77 et et COCO Cc dico-forme
226 et et COCO Cc dico-forme
270 et et COCO Cc dico-forme
372 et et COCO Cc dico-forme
392 et et COCO Cc dico-forme
1187 et et COCO Cc dico-forme
1244 et et COCO Cc dico-forme
1527 et et COCO Cc dico-forme
1591 et et COCO Cc dico-forme
1631 et et COCO Cc dico-forme
1724 et et COCO Cc dico-forme
2150 et et COCO Cc dico-forme
2205 et et COCO Cc dico-forme
2213 et et COCO Cc dico-forme
2337 et et COCO Cc dico-forme
2447 et et COCO Cc dico-forme
2556 et et COCO Cc dico-forme
2612 et et COCO Cc dico-forme
2813 et et COCO Cc dico-forme
2929 et et COCO Cc dico-forme
3306 et et COCO Cc dico-forme
3468 et et COCO Cc dico-forme
3922 et et COCO Cc dico-forme
4145 et et COCO Cc dico-forme
4254 et et COCO Cc dico-forme
4317 et et COCO Cc dico-forme
4348 et et COCO Cc dico-forme
4675 et et COCO Cc dico-forme
4831 et et COCO Cc dico-forme
5282 et et COCO Cc dico-forme
5438 et et COCO Cc dico-forme
5640 et et COCO Cc dico-forme
5789 et et COCO Cc dico-forme

Figure 127 : *Le Trameur* sélection sur dictionnaires (onglets FORME, LEMME, CATEGORIE)

Dans la première figure qui précède, les items surlignés vont être sélectionnés : l'ensemble de leurs positions respectives sont mémorisées (visibles dans le seconde figure) dans le *Gestionnaire de Sélection*.

21.2 Traitements d'une sélection dans le Gestionnaire de sélection

L'outil *Gestionnaire de Sélection* est visible dans l'onglet **SELECTION**. Le *Gestionnaire de Sélection* permet :

- De visualiser les items sélectionnés (onglet **Liste Sélections**)
- De lancer des opérations sur une sélection de positions disponibles dans le *Gestionnaire de Sélection*
- De rechercher des items en filtrant sur le niveau d'annotation n°1 (forme) (onglet **Recherche Forme**) : les items résultants peuvent ensuite être ajoutés au *Gestionnaire de Sélection*
- De rechercher des items en filtrant sur les différents niveaux d'annotation disponibles (onglet **Recherche Annotation(s)**) : les items résultants peuvent ensuite être ajoutés au *Gestionnaire de Sélection*
- De rechercher des segments parmi ceux disponibles (quel que soit le niveau d'annotation choisi pour le calcul des segments) (onglet **Recherche Segments**) : les segments résultants peuvent ensuite être ajoutés au *Gestionnaire de Sélection*
- De sélectionner des « zones » du ***Cadre*** (une ou plusieurs parties) pour y lancer des opérations textométriques (onglet **Sélections sur *Cadre***)

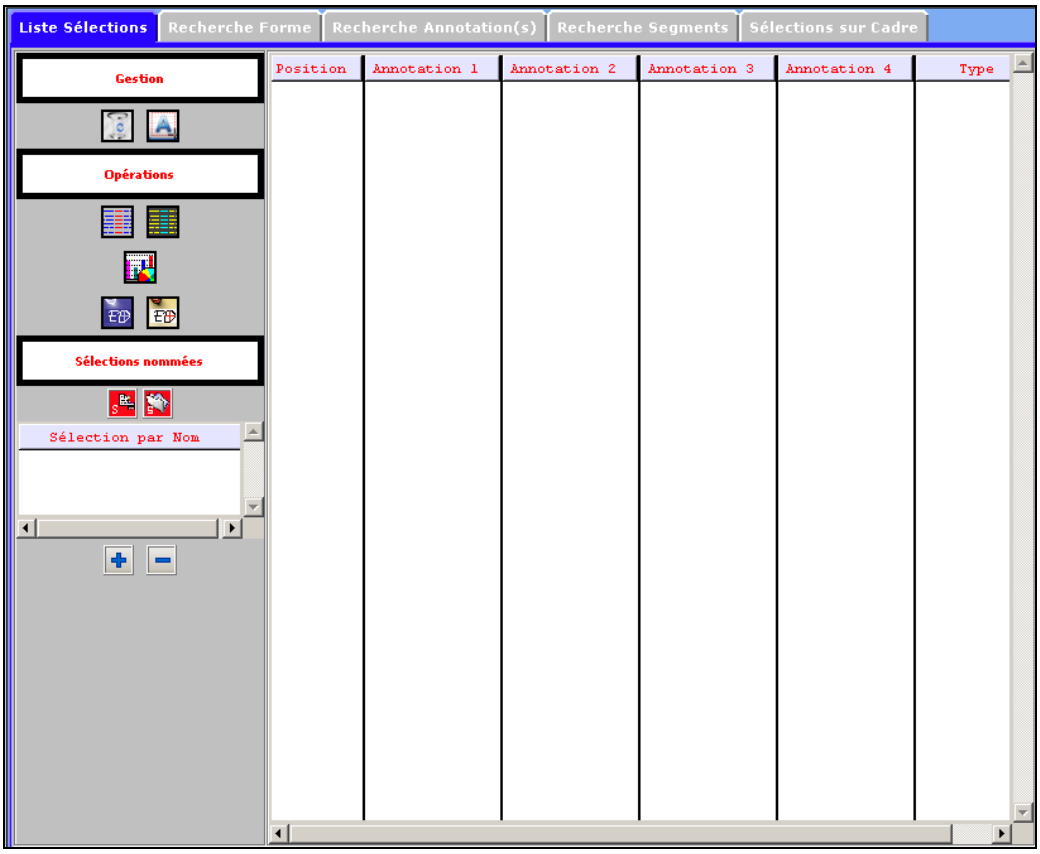


Figure 128 : *Le Trameur Gestionnaire de Sélection*

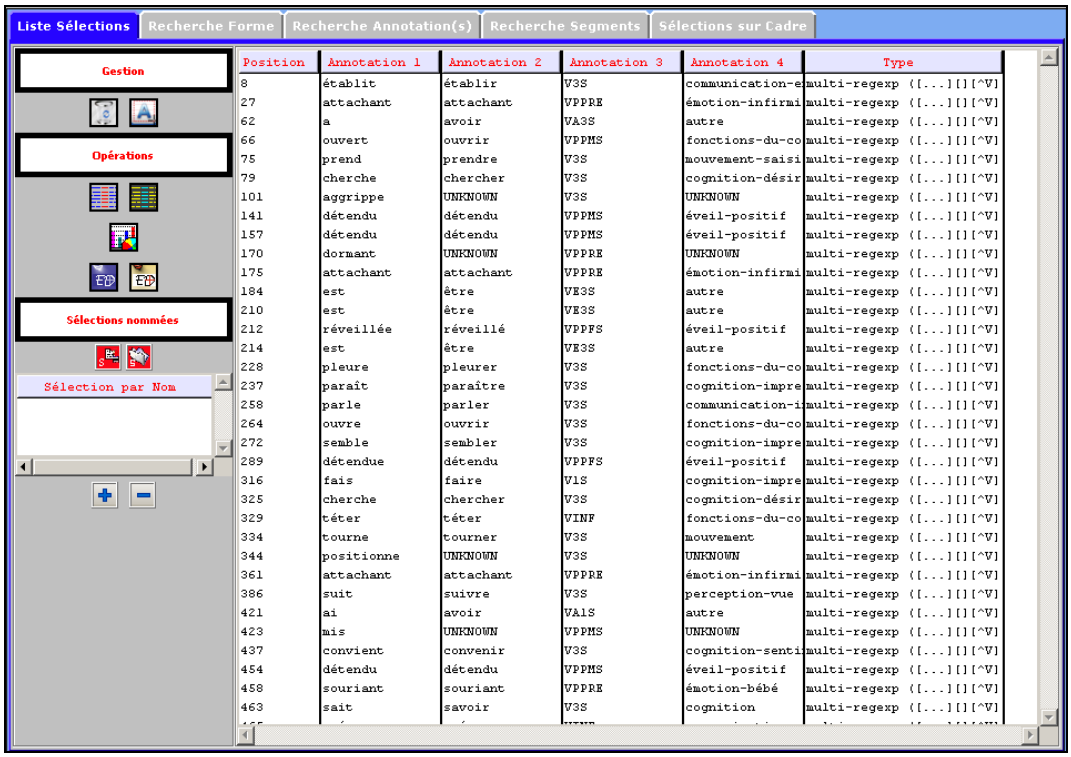


Figure 129 : *Le Trameur* *Gestionnaire de Sélection* (positions stockées dans le *Gestionnaire de Sélection*)

On peut ensuite agir sur les positions sélectionnées *via* les boutons disponibles dans la partie supérieure de la fenêtre du *Gestionnaire de Sélection*.

On peut supprimer toutes les sélections réalisées ou les sélectionner toutes :

On peut, après sélection des positions mémorisées, lancer des opérations sur ces positions : *concordance*, *ventilation*, *carte des sections*, *modification des annotations*, *insertion d'une nouvelle annotation*

21.3 Opérations sur sélections : concordance, ventilation, modification annotation, nouvelle annotation

21.3.1 Concordance

L'exemple ci-dessous montre le texte de travail avant chargement :

```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
<LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
<LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
<LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
<LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut.
<LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
<LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
<LIGNE="VERS11">Nature, berce-le chaudement : il a froid.
<LIGNE="VERS12">Les parfums ne font pas frissonner sa narine ;
<LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
<LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit.
<STRUCTURE="AUTEUR">Arthur Rimbaud
```

La première figure qui suit montre l'ensemble des positions sélectionnées (contour jaune) dans l'éditeur de la ***Trame*** :

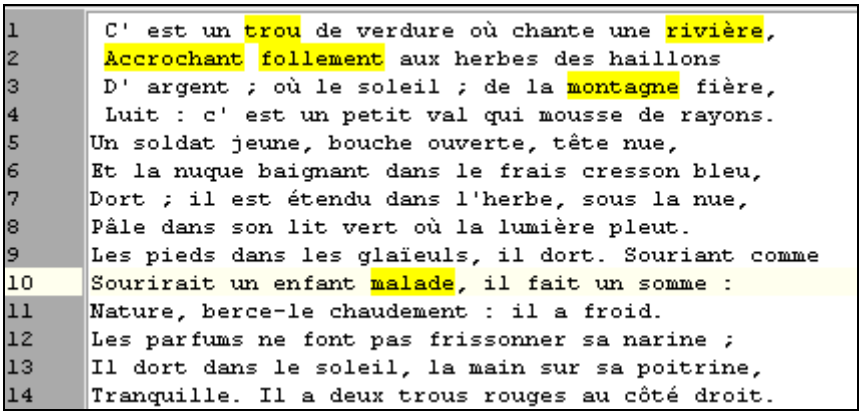

1 C' est un trou de verdure où chante une rivière,
2 Accrochant follement aux herbes des haillons
3 D' argent ; où le soleil ; de la montagne fière,
4 Luit : c' est un petit val qui mousse de rayons.
5 Un soldat jeune, bouche ouverte, tête nue,
6 Et la nuque baignant dans le frais cresson bleu,
7 Dort ; il est étendu dans l'herbe, sous la nue,
8 Pâle dans son lit vert où la lumière pleut.
9 Les pieds dans les glaïeuls, il dort. Souriant comme
10 Sourirait un enfant malade, il fait un somme :
11 Nature, berce-le chaudement : il a froid.
12 Les parfums ne font pas frissonner sa narine ;
13 Il dort dans le soleil, la main sur sa poitrine,
14 Tranquille. Il a deux trous rouges au côté droit.

Figure 130 : *Le Trameur* sélection directe de position dans l'éditeur

La figure suivante montre l'ensemble des positions mémorisées dans le *Gestionnaire de Sélection :*

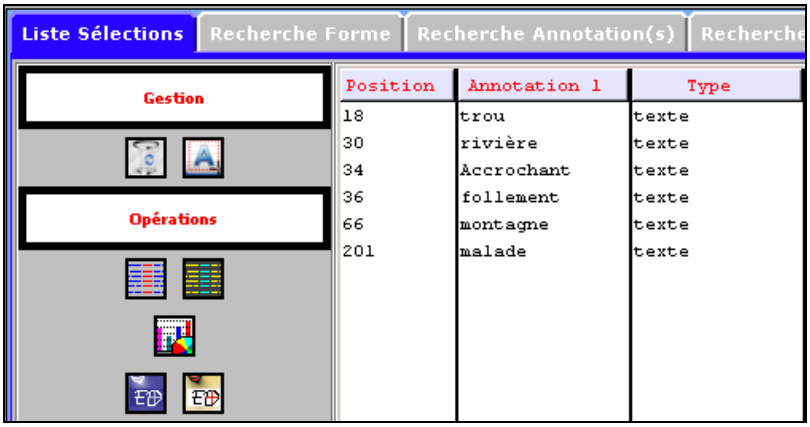


Figure 131 : *Le Trameur Gestionnaire de Sélection* : les positions enregistrées

La figure suivante montre la sélection réalisée parmi l'ensemble des positions mémorisées (surlignées en noir) dans le *Gestionnaire de Sélection* :

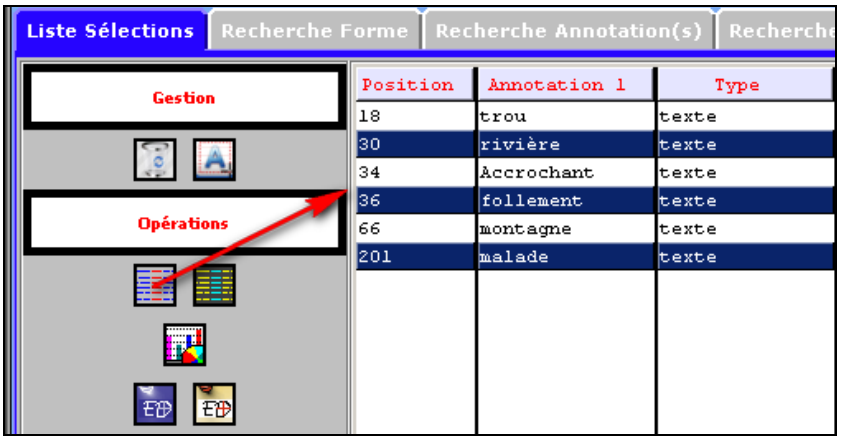


Figure 132 : *Le Trameur* *Gestionnaire de Sélection* : sélection parmi les positions enregistrées

Enfin la figure suivante montre le résultat de la concordance réalisée sur les positions sélectionnées :

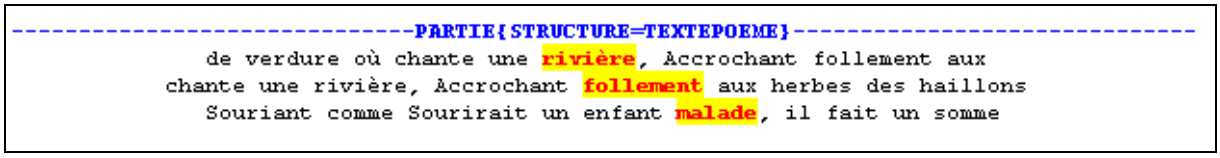

----------PARTIE{STRUCTURE=TEXTEPOEME}----------
de verdure où chante une rivière, Accrochant follement aux
chante une rivière, Accrochant follement aux herbes des haillons
Souriant comme Sourirait un enfant malade, il fait un somme

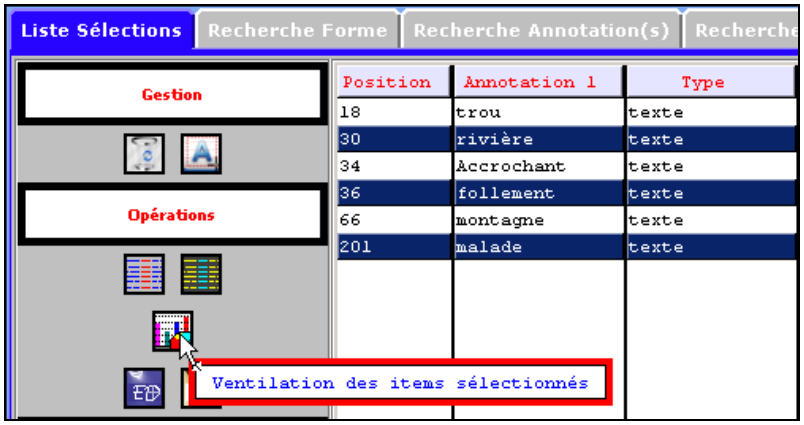
Figure 133 : *Le Trameur* concordance d'une sélection de positions

La concordance réalisée est la concordance des positions sélectionnées :

- position 30 (rivière)
- position 36 (follement)
- position 201 (malade)

21.3.2 Ventilation

En reprenant la même sélection, la figure suivante montre le résultat de la ventilation réalisée sur les positions sélectionnées sur la partition VERS.



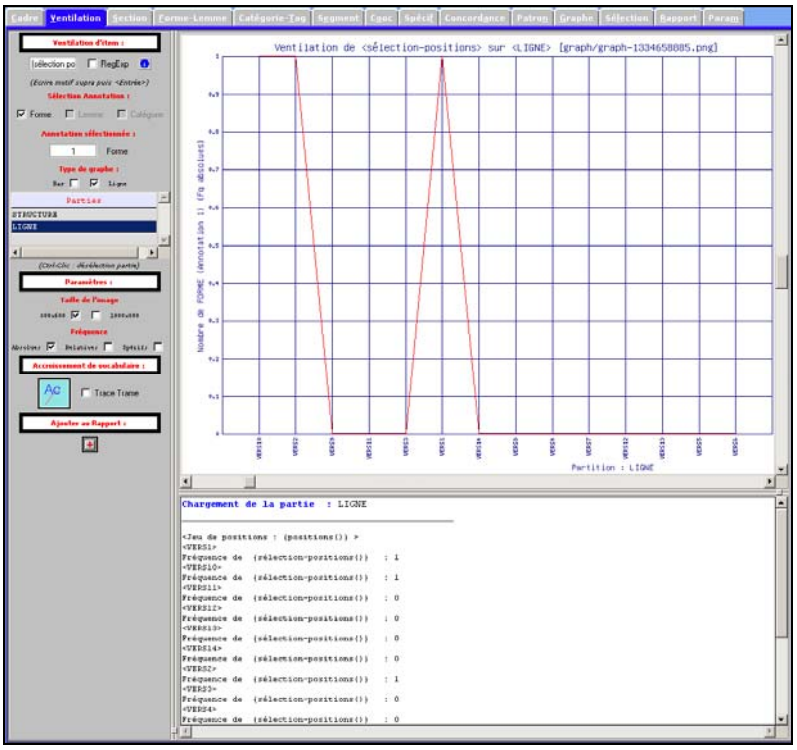


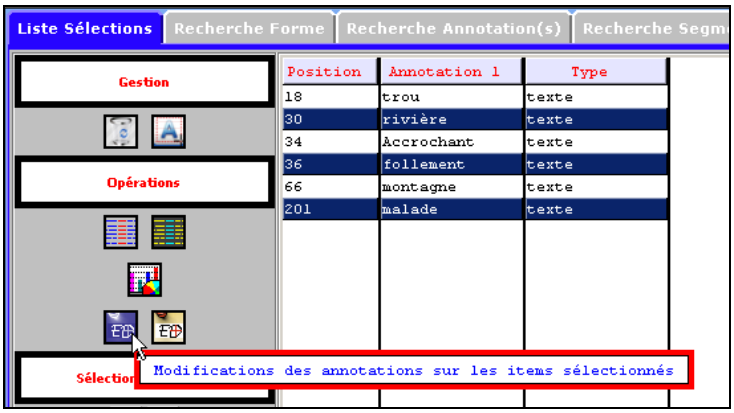
Figure 134 : *Le Trameur* ventilation d'une sélection de positions

La ventilation réalisée est la ventilation des formes associées aux positions sélectionnées :

- position 30 (rivière) : 1 occurrence de cette forme dans la partie VERS1
- position 36 (follement) : 1 occurrence de cette forme dans la partie VERS2
- position 201 (malade) : 2 occurrences de cette forme dans la partie VERS10

21.3.3 Modification des annotations sur les items sélectionnés

Toujours en reprenant la même sélection, la figure suivante montre comment modifier les annotations de tous les items sélectionnés :



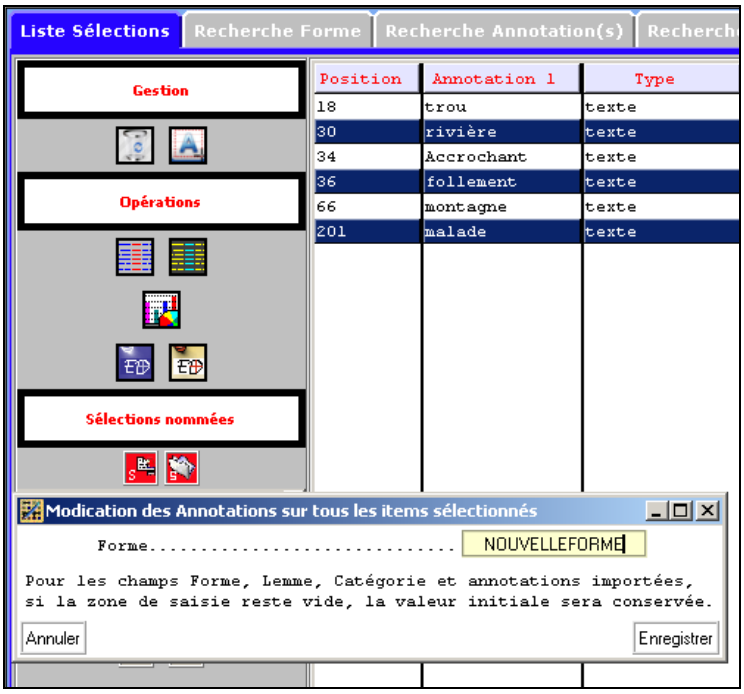


Figure 135 : *Le Trameur* modification des annotations d'une sélection de positions

Une zone de saisie permet de mettre à jour les annotations de tous les items correspondant aux positions sélectionnées. Pour modifier une annotation, il suffit de saisir la nouvelle annotation dans la zone de saisie idoine : les zones de saisie « vides » permettent de conserver les annotations initiales.

Après modification, le *Gestionnaire de Sélection* est mis à jour :

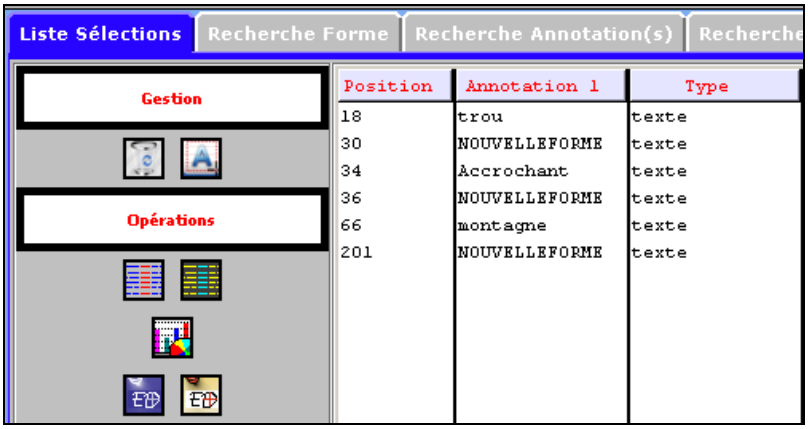


Figure 136 : *Le Trameur* modification des annotations d'une sélection de positions

Ainsi que le dictionnaire des formes :

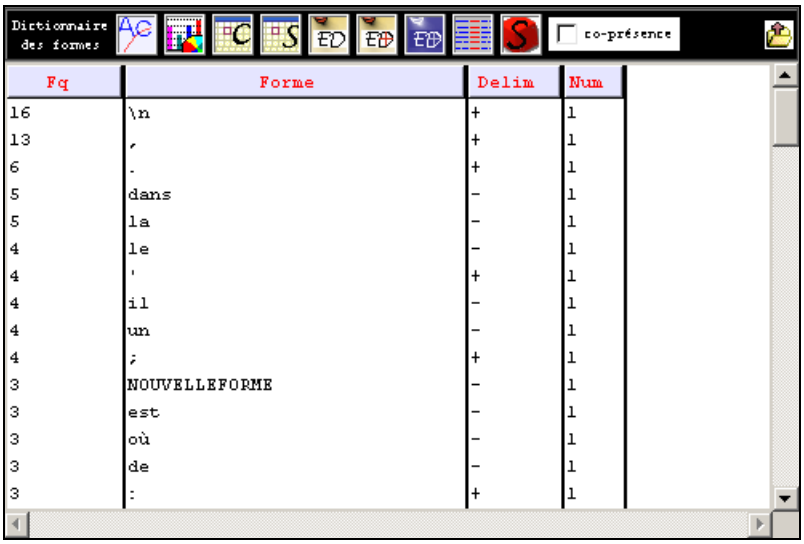


Figure 137 : *Le Trameur* modification des annotations d'une sélection de positions

21.3.4 Ajout d'une nouvelle annotation sur les items sélectionnés

Toujours en reprenant la même sélection, la figure suivante montre comment modifier les annotations de tous les items sélectionnés :

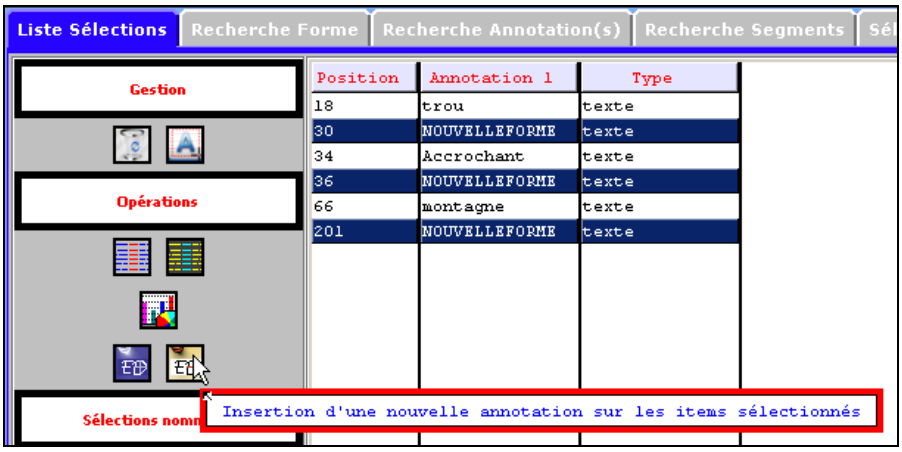


Figure 138 : *Le Trameur* insertion d'une nouvelle annotation sur une sélection de positions

Comme dans l'étape précédente, après avoir sélectionné des items dans le *Gestionnaire de Sélection*, une zone de saisie permet de définir une nouvelle couche d'annotation sur tous les items correspondant aux positions sélectionnées.

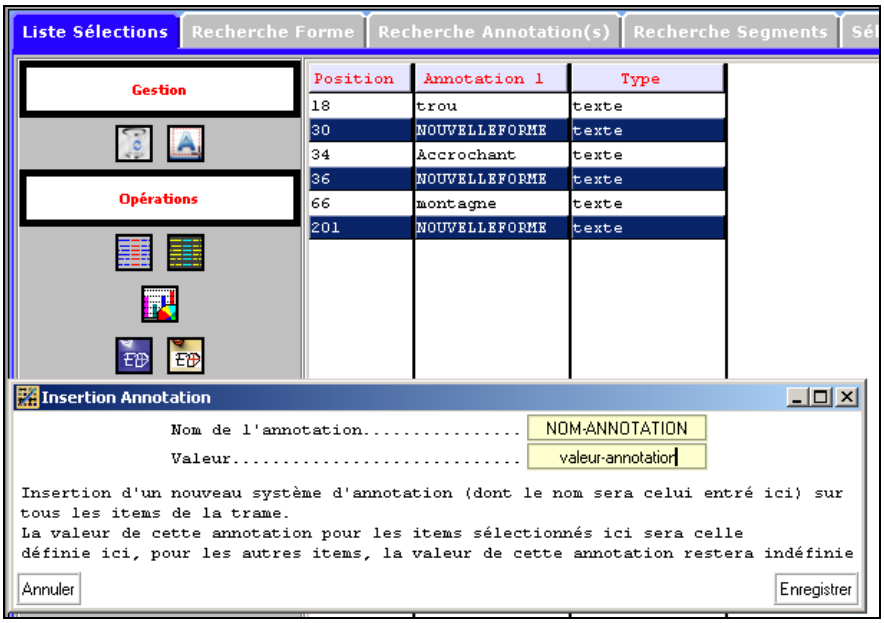


Figure 139 : *Le Trameur* insertion d'une nouvelle annotation sur une sélection de positions

Pour insérer une annotation, il suffit de saisir le nom de la nouvelle annotation et sa valeur dans la zone de saisie idoine. Les autres items de la ***Trame*** auront aussi cette nouvelle couche d'annotation mais avec une valeur indéfinie (« unknown »).

Après modification, le *Gestionnaire de Sélection* est mis à jour :

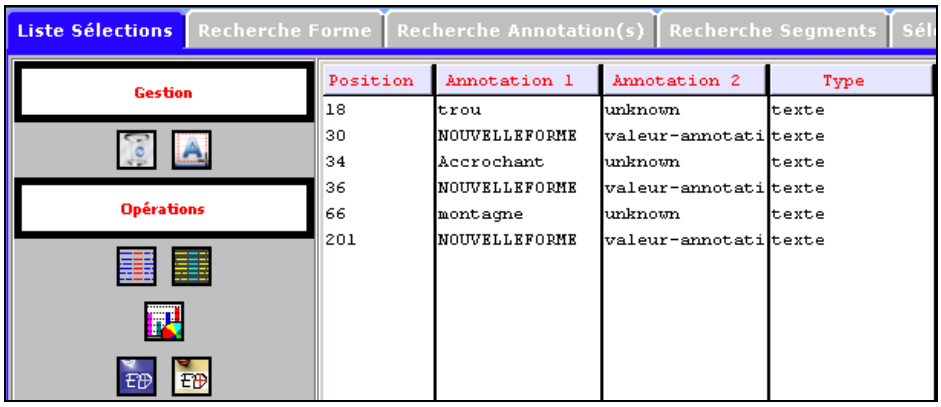


Figure 140 : *Le Trameur* insertion d'une nouvelle annotation sur une sélection de positions

Ainsi que le dictionnaire des annotations :

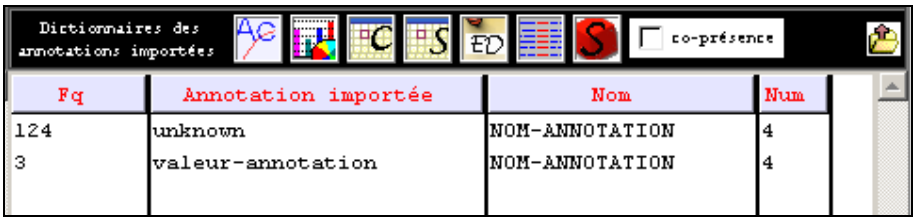


Figure 141 : *Le Trameur* insertion d'une nouvelle annotation sur une sélection de positions

21.4 Méthodes de génération de sélection via le Gestionnaire

21.4.1 Génération et sélection d'une liste de formes

L'onglet **Recherche Forme** du *Gestionnaire de Sélection* permet de lancer la recherche d'une liste de formes possédant une propriété décrite par un motif exprimé sous la forme d'une expression régulière. Dans la figure suivante, on recherche une liste de formes possédant la propriété : `\bd` (forme commençant par d).

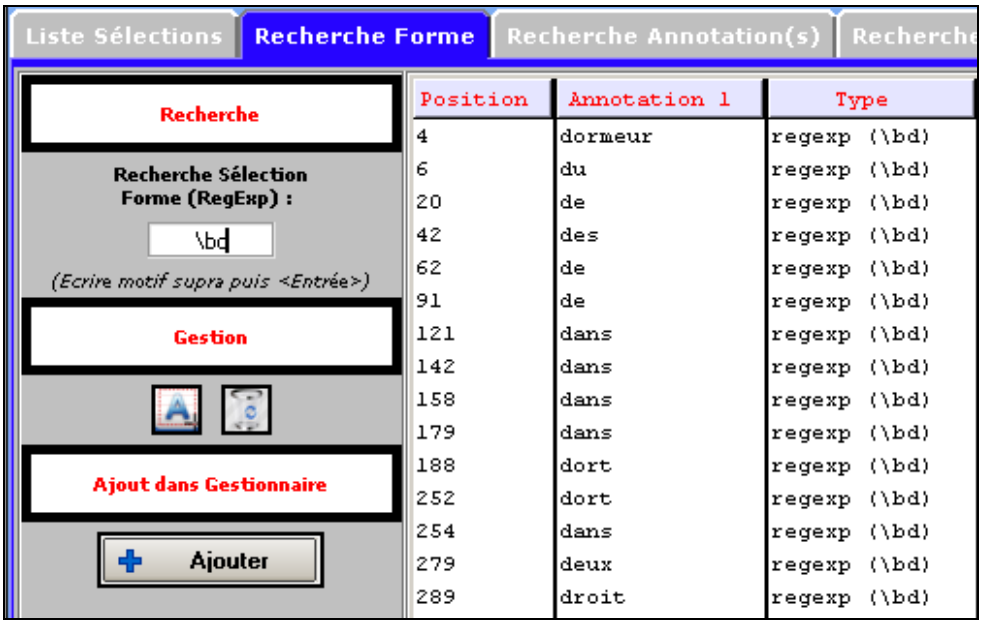


Figure 142 : *Le Trameur* génération d'une liste de formes à sélectionner

On peut ensuite choisir les positions à sélectionner *i.e.* choisir celles qui seront ajoutées laux sélections en cours dans l'onglet **Liste Sélections**.

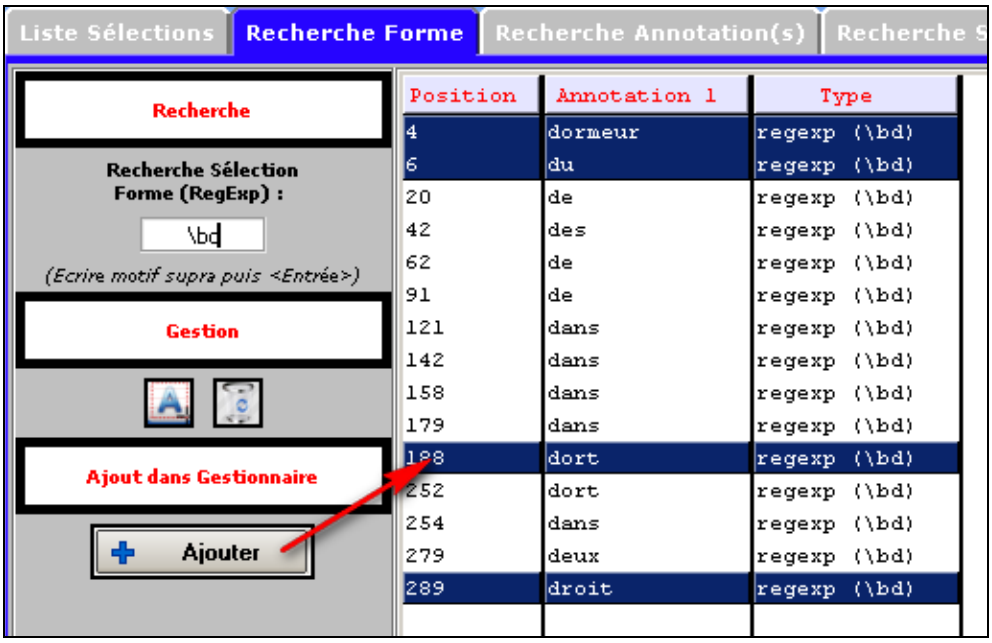


Figure 143 : *Le Trameur* sélection d'une liste de formes

21.4.2 Génération et sélection d'une liste de segments

L'onglet **Recherche Segments** du *Gestionnaire de Sélection* permet de lancer la recherche d'une liste de segments contenant un motif exprimé sous la forme d'une expression régulière. Dans la figure suivante, on recherche une liste de segments contenant la forme « le ».

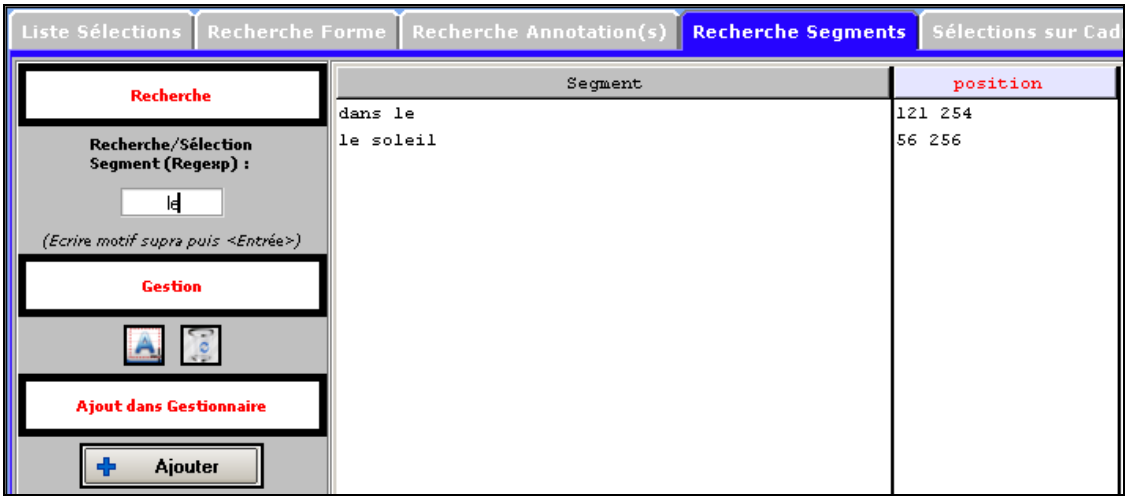


Figure 144 : *Le Trameur* génération d'une liste de segments à sélectionner

Comme précédemment, on peut ensuite choisir les segments à sélectionner et les ajouter aux sélections en cours dans l'onglet **Liste Sélections**. Cet ajout consiste à mémoriser la position initiale du segment visé : les positions ajoutées sont celles associées à la première forme du segment visé (dans l'exemple ci-dessus : le segment « dans le » est présent à la position 121 et 254 de la ***Trame*** (positions respectives des 2 formes « dans » amorçant le segment), ces positions mémorisées seront aussi associées à la longueur du segment ajouté (dans l'exemple ci-dessus, on mémorise donc les positions 121 et 254 et la longueur 2).

21.4.3 Génération et sélection multicritère

Après chargement d'une base textométrique, chaque item de la ***Trame*** est associé à un liste d'annotation, cette liste peut être enrichie pour ajout dynamique d'une nouvelle couche d'annotation (cf *Ajout/modification d'annonation sur la* ***Trame***). L'onglet **Recherche d'annotation(s)** permet de sélectionner des items de la ***Trame*** en paramétrant une recherche multicritère sur les différentes annotations disponibles pour chaque item.

L'interface de cet onglet est reconstruite dynamiquement dès qu'une annotation supplémentaire est intégrée dans la base en cours de traitement.

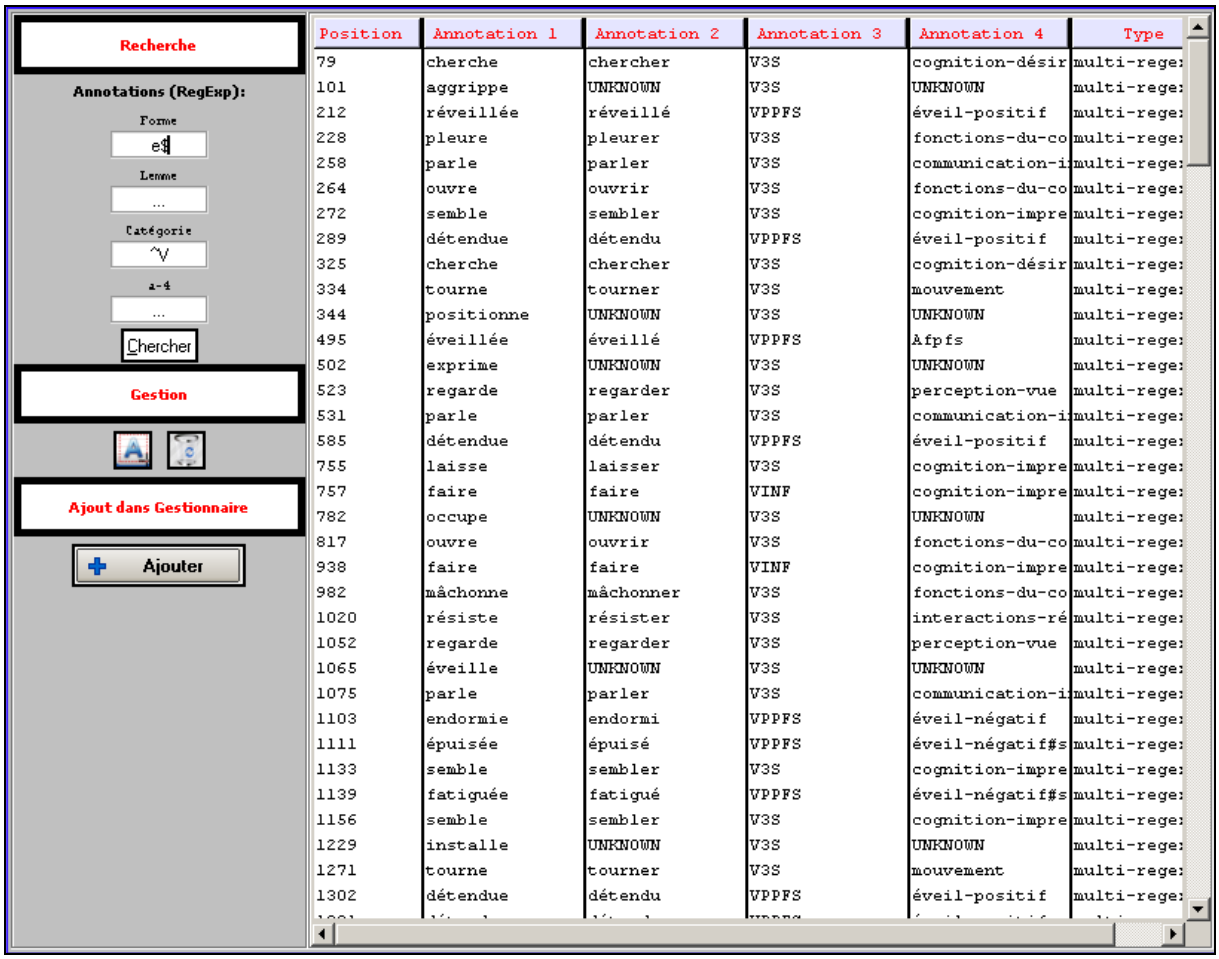

Recherche
Annotations (RegExp):
Forme
e$
Lemme
...
Catégorie
^V
a-4
...
Chercher
Gestion
Ajout dans Gestionnaire
Ajouter
Position | Annotation 1 | Annotation 2 | Annotation 3 | Annotation 4 | Type
79 cherche chercher V3S cognition-désir multi-rege
101 aggrippe UNKNOWN V3S UNKNOWN multi-rege
212 réveillée réveillé VPPFS éveil-positif multi-rege
228 pleure pleurer V3S fonctions-du-co multi-rege
258 parle parler V3S communication-i multi-rege
264 ouvre ouvrir V3S fonctions-du-co multi-rege
272 semble sembler V3S cognition-impre multi-rege
289 détendue détendu VPPFS éveil-positif multi-rege
325 cherche chercher V3S cognition-désir multi-rege
334 tourne tourner V3S mouvement multi-rege
344 positionne UNKNOWN V3S UNKNOWN multi-rege
495 éveillée éveillé VPPFS Afpfs multi-rege
502 exprime UNKNOWN V3S UNKNOWN multi-rege
523 regarde regarder V3S perception-vue multi-rege
531 parle parler V3S communication-i multi-rege
585 détendue détendu VPPFS éveil-positif multi-rege
755 laisse laisser V3S cognition-impre multi-rege
757 faire faire VINF cognition-impre multi-rege
782 occupe UNKNOWN V3S UNKNOWN multi-rege
817 ouvre ouvrir V3S fonctions-du-co multi-rege
938 faire faire VINF cognition-impre multi-rege
982 mâchonne mâchonner V3S fonctions-du-co multi-rege
1020 résiste résister V3S interactions-ré multi-rege
1052 regarde regarder V3S perception-vue multi-rege
1065 éveille UNKNOWN V3S UNKNOWN multi-rege
1075 parle parler V3S communication-i multi-rege
1103 endormie endormi VPPFS éveil-négatif multi-rege
1111 épuisée épuisé VPPFS éveil-négatif#s multi-rege
1133 semble sembler V3S cognition-impre multi-rege
1139 fatiguée fatigué VPPFS éveil-négatif#s multi-rege
1156 semble sembler V3S cognition-impre multi-rege
1229 installe UNKNOWN V3S UNKNOWN multi-rege
1271 tourne tourner V3S mouvement multi-rege
1302 détendue détendu VPPFS éveil-positif multi-rege

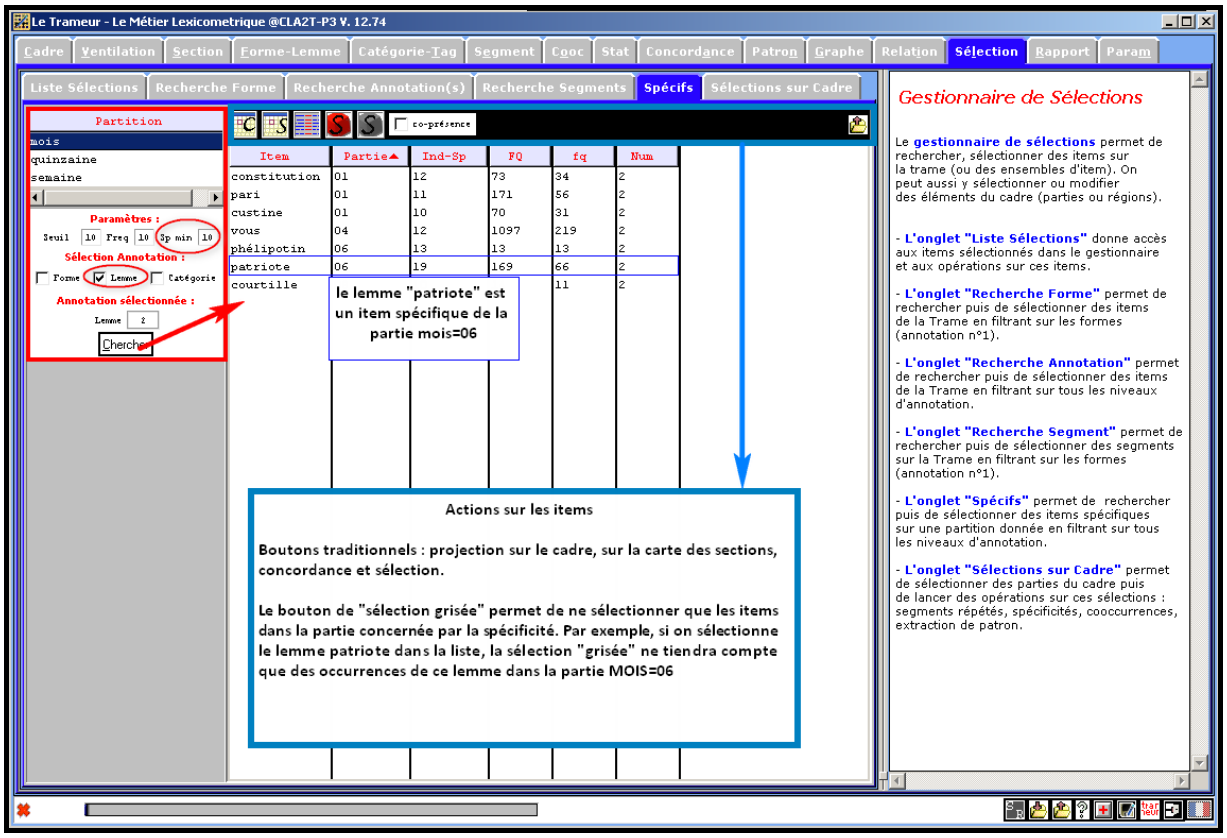
Figure 145 : *Le Trameur* génération d'une liste d'items par équation de recherche

Dans l'exemple précédent, la recherche définie :

permet de sélectionner les items de la ***Trame*** ayant un niveau d'annotation n°1 (Forme) se terminant par la lettre e et un niveau d'annotation n°3 (Catégorie) commençant par la lettre V, les 2 autres niveaux d'annotation étant indifférents. La liste produite illustre les items de la ***Trame*** possédant ces propriétés.

21.4.4 Génération d'une liste d'items spécifiques

L'onglet **Spécifs** du *Gestionnaire de Sélection* permet de lancer la recherche d'une liste d'items spécifiques sur une partition donnée. Dans la figure suivante, on recherche des lemmes spécifiques (ici un indice de spécificité supérieur à 10) sur la partition MOIS de la base chargée.


Le Trameur - Le Métier Lexicometrique @CLA2T-P3 V. 12.74
Cadre
Ventilation
Section
Forme-Lemme
Catégorie-Tag
Segment
Cooc
Stat
Concordance
Patron
Graphe
Relation
Sélection
Rapport
Param
Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Spécifs
Sélections sur Cadre
Partition
mois
quinzaine
semaine
Paramètres :
Seuil 10 Freq 10 Sp min 10
Sélection Annotation :
Forme
Lemme
Catégorie
Annotation sélectionnée :
Lemme 2
Chercher
co-présence
Item
Partie
Ind-Sp
FQ
fq
Num
constitution 01 12 73 34 2
pari 01 11 171 56 2
custine 01 10 70 31 2
vous 04 12 1097 219 2
phélipotin 06 13 13 13 2
patriote 06 19 169 66 2
courtille 11 2
le lemme "patriote" est un item spécifique de la partie mois=06
Actions sur les items
Boutons traditionnels : projection sur le cadre, sur la carte des sections, concordance et sélection.
Le bouton de "sélection grisée" permet de ne sélectionner que les items dans la partie concernée par la spécificité. Par exemple, si on sélectionne le lemme patriote dans la liste, la sélection "grisée" ne tiendra compte que des occurrences de ce lemme dans la partie MOIS=06
Gestionnaire de Sélections
Le gestionnaire de sélections permet de rechercher, sélectionner des items sur la trame (ou des ensembles d'item). On peut aussi y sélectionner ou modifier des éléments du cadre (parties ou régions).
- L'onglet "Liste Sélections" donne accès aux items sélectionnés dans le gestionnaire et aux opérations sur ces items.
- L'onglet "Recherche Forme" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur les formes (annotation n°1).
- L'onglet "Recherche Annotation" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur tous les niveaux d'annotation.
- L'onglet "Recherche Segment" permet de rechercher puis de sélectionner des segments sur la Trame en filtrant sur les formes (annotation n°1).
- L'onglet "Spécifs" permet de rechercher puis de sélectionner des items spécifiques sur une partition donnée en filtrant sur tous les niveaux d'annotation.
- L'onglet "Sélections sur Cadre" permet de sélectionner des parties du cadre puis de lancer des opérations sur ces sélections : segments répétés, spécificités, cooccurrences, extraction de patron.

Le résultat donne à voir les items (lemmes) spécifiques sur les parties de la partition visée. Ces éléments peuvent ensuite être utilisés pour des opérations déjà présentées : ventilation, concordance, carte des sections, sélection.

Le bouton de sélection « grisé » permet d'ajouter au *Gestionnaire de Sélection* uniquement les occurrences de la partie visée.

21.5 Sélection de parties (comme ensembles de sélection de positions sur la Trame)

Le ***Cadre*** d'une base textométrique est défini comme un système permettant des accès aux textes, il regroupe des systèmes de parties (partition recouvrante ou non de la surface textuelle). **On peut de fait assimiler ces accès aux textes à des sélections de positions sur la *Trame***.
Si on considère le texte de travail suivant présenté dans 2 formats possibles en entrée du ***Trameur*** :

```
<STRUCTURE="TITRE">Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C'est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D'argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c'est un petit val qui mousse de rayons.
```

```
<?xml version="1.0" encoding="iso-8859-1"?>
<poeme>
<structure niveau="TITRE">Le dormeur du val</structure>
<structure niveau="TEXTEPOEME">
<ligne type="VERS1"> C'est un trou de verdure où chante une rivière,</ligne>
<ligne type="VERS2"> Accrochant follement aux herbes des haillons</ligne>
<ligne type="VERS3"> D'argent ; où le soleil ; de la montagne fière,</ligne>
<ligne type="VERS4"> Luit : c'est un petit val qui mousse de rayons.</ligne>
</structure>
</poeme>
```

Une segmentation *choisie* de la partie textuelle de ce texte (pour ces 2 formats) peut conduire à une ***Trame*** décrite ainsi :

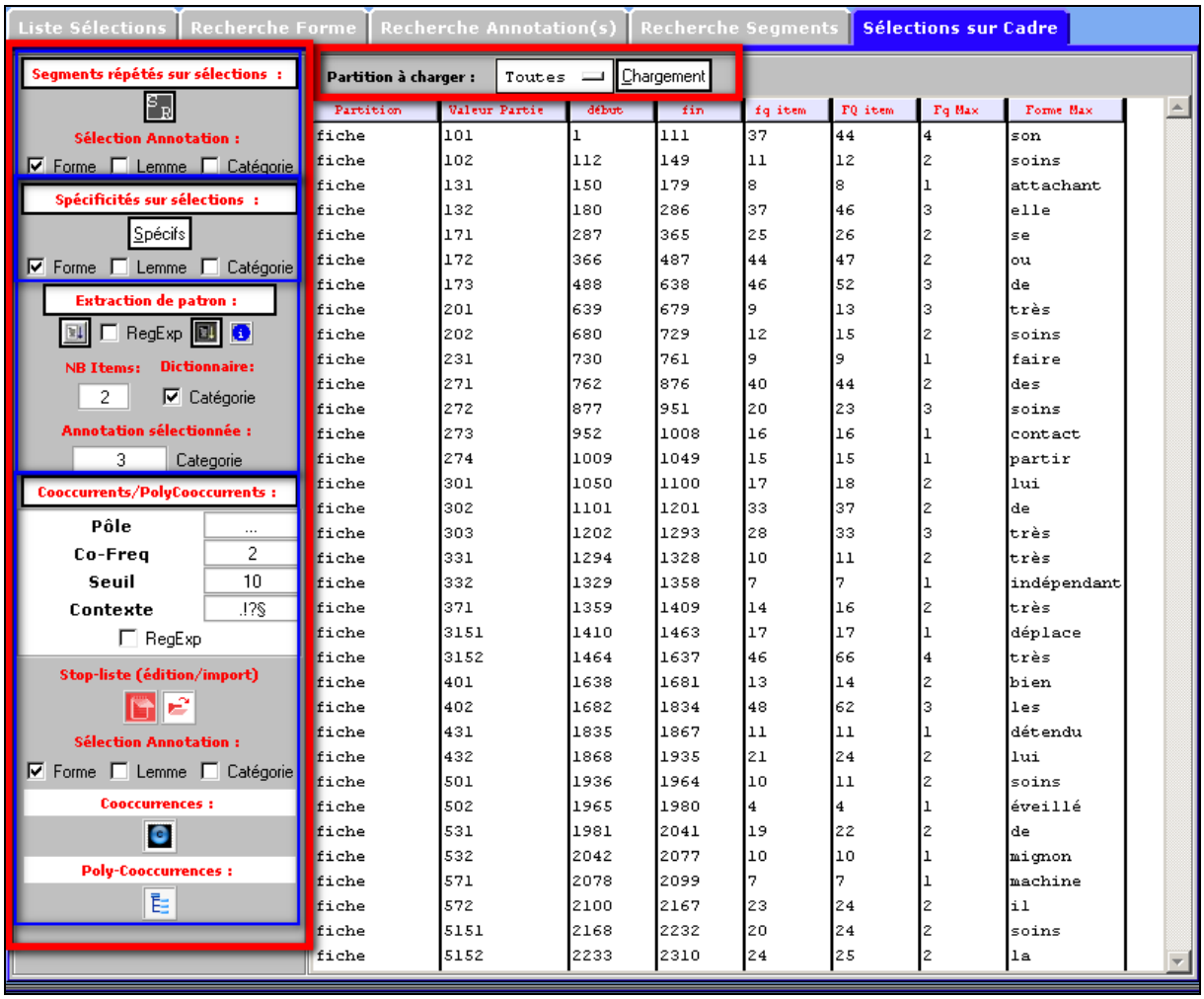
Forme	**Le**		**dormeur**		**du**		**val**	**\n**	**…**	**rayons**	**.**
Positions	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**…**	**pos-1**	**pos**

(pos désignant la dernière position de la ***Trame***)
Pour le texte dans son premier format, les partitions disponibles peuvent être décrites ainsi :
Partition STRUCTURE
La partie STRUCTURE_TITRE couvre la ***Trame*** de la position 1 à la position 8
La partie STRUCTURE_TEXTEPOEME couvre la ***Trame*** de la position 9 à la position *pos*
Partition LIGNE
La partie LIGNE_VERS1 couvre la ***Trame*** de la position 9 à la position …
Etc.

Pour le texte dans son second format, les partitions disponibles peuvent être décrites ainsi :
Partition poeme
La partie poeme couvre la ***Trame*** de la position 1 à la position pos
Partition structure
La partie structure_TITRE couvre la ***Trame*** de la position 1 à la position 8
La partie structure_TEXTEPOEME couvre la ***Trame*** de la position 9 à la position *pos*
Partition ligne
La partie ligne_VERS1 couvre la ***Trame*** de la position 9 à la position …
Etc.

Chacun des parties décrites est ainsi une sélection de positions sur la ***Trame*** couvrant les items compris entre la position initiale et la position finale sur la ***Trame*** décrites pour chaque partie.

L'outil *Gestionnaire de Sélection* dispose d'un onglet permettant de sélectionner des zones sur la ***Trame*** en s'appuyant sur les parties c'est-à-dire sur les sélections de positions définies par le ***Cadre*** de la base textométrique.



Partition	Valeur Partie	début	fin	fq item	FQ item	Fq Max	Forme Max
fiche	101	1	111	37	44	4	son
fiche	102	112	149	11	12	2	soins
fiche	131	150	179	8	8	1	attachant
fiche	132	180	286	37	46	3	elle
fiche	171	287	365	25	26	2	se
fiche	172	366	487	44	47	2	ou
fiche	173	488	638	46	52	3	de
fiche	201	639	679	9	13	3	très
fiche	202	680	729	12	15	2	soins
fiche	231	730	761	9	9	1	faire
fiche	271	762	876	40	44	2	des
fiche	272	877	951	20	23	3	soins
fiche	273	952	1008	16	16	1	contact
fiche	274	1009	1049	15	15	1	partir
fiche	301	1050	1100	17	18	2	lui
fiche	302	1101	1201	33	37	2	de
fiche	303	1202	1293	28	33	3	très
fiche	331	1294	1328	10	11	2	très
fiche	332	1329	1358	7	7	1	indépendant
fiche	371	1359	1409	14	16	2	très
fiche	3151	1410	1463	17	17	1	déplace
fiche	3152	1464	1637	46	66	4	très
fiche	401	1638	1681	13	14	2	bien
fiche	402	1682	1834	48	62	3	les
fiche	431	1835	1867	11	11	1	détendu
fiche	432	1868	1935	21	24	2	lui
fiche	501	1936	1964	10	11	2	soins
fiche	502	1965	1980	4	4	1	éveillé
fiche	531	1981	2041	19	22	2	de
fiche	532	2042	2077	10	10	1	mignon
fiche	571	2078	2099	7	7	1	machine
fiche	572	2100	2167	23	24	2	il
fiche	5151	2168	2232	20	24	2	soins
fiche	5152	2233	2310	24	25	2	la

Figure 146 : *Le Trameur* sélection de parties

Une ou plusieurs parties étant sélectionnées, 4 types d'opérations sont disponibles sur les parties sélectionnées :

- calcul des segments
- calcul de spécicifité
- extraction de patron
- calcul de cooccurrents / polycooccurrents

Les figures suivantes illustrent ces différents modules. Dans tous les cas, après avoir sélectionné une ou plusieurs parties, les calculs invoqués se font sur la sélection visée.

21.5.1 Calcul de segments

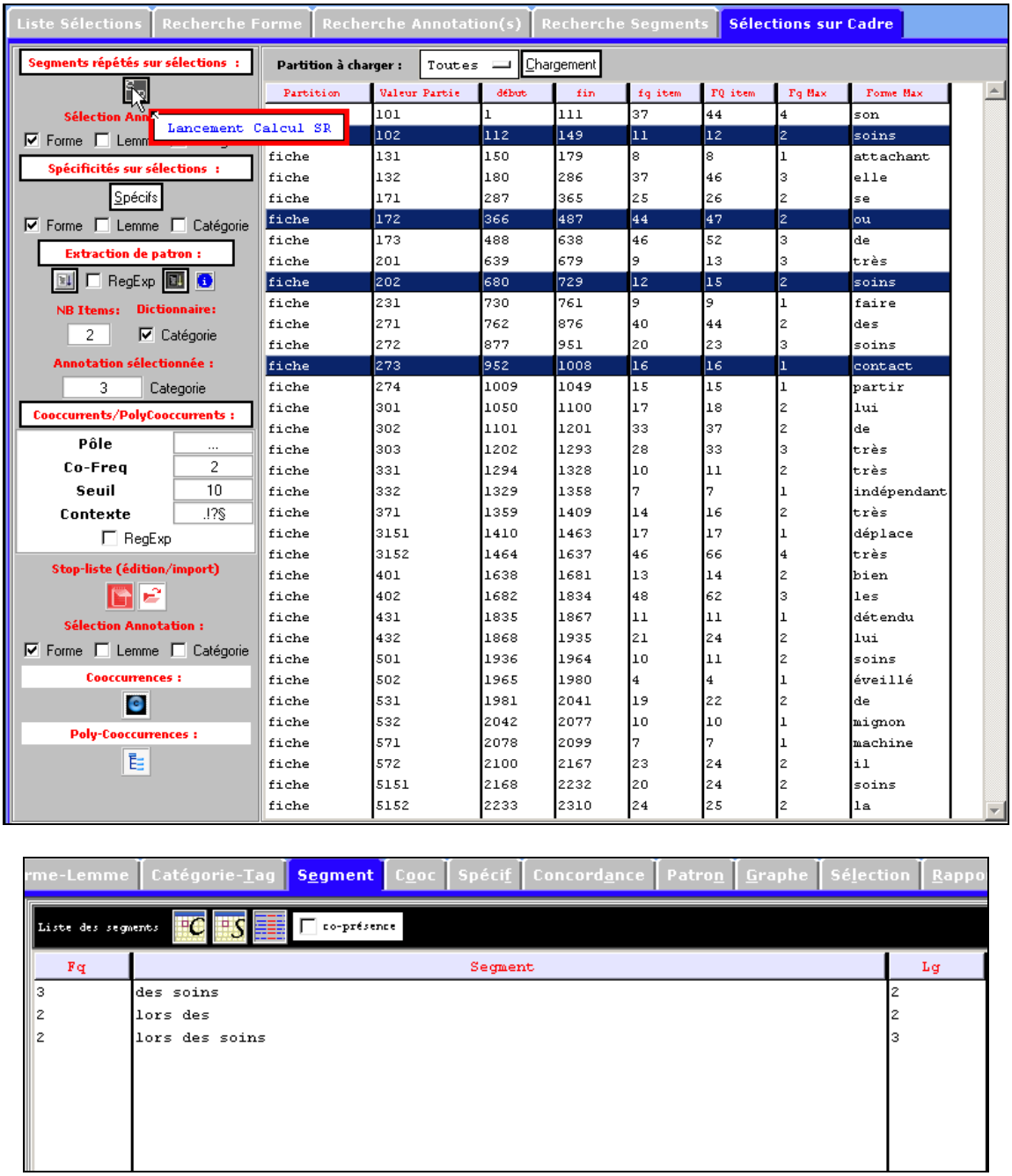

Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Sélections sur Cadre
Segments répétés sur sélections :
Lancement Calcul SR
Spécificités sur sélections :
Spécifs
Extraction de patron :
RegExp
NB Items:
Dictionnaire:
Catégorie
Annotation sélectionnée :
Categorie
Cooccurrents/PolyCooccurrents :
Pôle
Co-Freq
Seuil
Contexte
Stop-liste (édition/import)
Sélection Annotation :
Cooccurrences :
Poly-Cooccurrences :
Partition à charger :
Toutes
Chargement
Partition
Valeur Partie
début
fin
fq item
FQ item
Fq Max
Forme Max
fiche 101 1 111 37 44 4 son
fiche 102 112 149 11 12 2 soins
fiche 131 150 179 8 8 1 attachant
fiche 132 180 286 37 46 3 elle
fiche 171 287 365 25 26 2 se
fiche 172 366 487 44 47 2 ou
fiche 173 488 638 46 52 3 de
fiche 201 639 679 9 13 3 très
fiche 202 680 729 12 15 2 soins
fiche 231 730 761 9 9 1 faire
fiche 271 762 876 40 44 2 des
fiche 272 877 951 20 23 3 soins
fiche 273 952 1008 16 16 1 contact
fiche 274 1009 1049 15 15 1 partir
fiche 301 1050 1100 17 18 2 lui
fiche 302 1101 1201 33 37 2 de
fiche 303 1202 1293 28 33 3 très
fiche 331 1294 1328 10 11 2 très
fiche 332 1329 1358 7 7 1 indépendant
fiche 371 1359 1409 14 16 2 très
fiche 3151 1410 1463 17 17 1 déplace
fiche 3152 1464 1637 46 66 4 très
fiche 401 1638 1681 13 14 2 bien
fiche 402 1682 1834 48 62 3 les
fiche 431 1835 1867 11 11 1 détendu
fiche 432 1868 1935 21 24 2 lui
fiche 501 1936 1964 10 11 2 soins
fiche 502 1965 1980 4 4 1 éveillé
fiche 531 1981 2041 19 22 2 de
fiche 532 2042 2077 10 10 1 mignon
fiche 571 2078 2099 7 7 1 machine
fiche 572 2100 2167 23 24 2 il
fiche 5151 2168 2232 20 24 2 soins
fiche 5152 2233 2310 24 25 2 la

Figure 147 : *Le Trameur* calcul de segment sur une sélection de parties

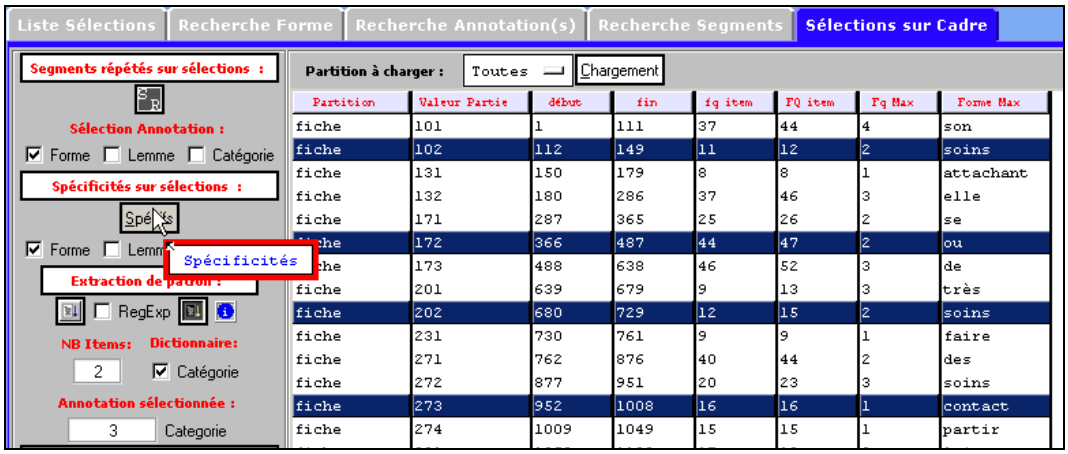
On commence par sélectionner des parties (ici 4 parties) puis on déclenche le calcul des SR avec le bouton idoine. Les segments résultants ne sont en fait disponibles que dans 2 des parties sélectionnées : une concordance des segments construits donne à voir leurs réalisations sur la ***Trame*** :

```
------------------------------PARTIE{fiche=102}------------------------------
                    . calme en dehors des soins . détendu au
------------------------------PARTIE{fiche=202}------------------------------
                  . sensible au toucher lors des soins . grimaces
                  . sensible au toucher lors des soins . grimaces
                sensible au toucher lors des soins . grimaces lors
                      des soins . grimaces lors des soins .
                      des soins . grimaces lors des soins .
                      soins . grimaces lors des soins . très
```

Figure 148 : *Le Trameur* calcul de segment sur une sélection de parties

21.5.2 Calcul de spécificité

Après avoir sélectionné des parties, on lance le calcul des spécificités :



Partition	Valeur Partie	début	fin	fq item	FQ item	Fq Max	Forme Max
fiche	101	1	111	37	44	4	son
fiche	102	112	149	11	12	2	soins
fiche	131	150	179	8	8	1	attachant
fiche	132	180	286	37	46	3	elle
fiche	171	287	365	25	26	2	se
fiche	172	366	487	44	47	2	ou
fiche	173	488	638	46	52	3	de
fiche	201	639	679	9	13	3	très
fiche	202	680	729	12	15	2	soins
fiche	231	730	761	9	9	1	faire
fiche	271	762	876	40	44	2	des
fiche	272	877	951	20	23	3	soins
fiche	273	952	1008	16	16	1	contact
fiche	274	1009	1049	15	15	1	partir

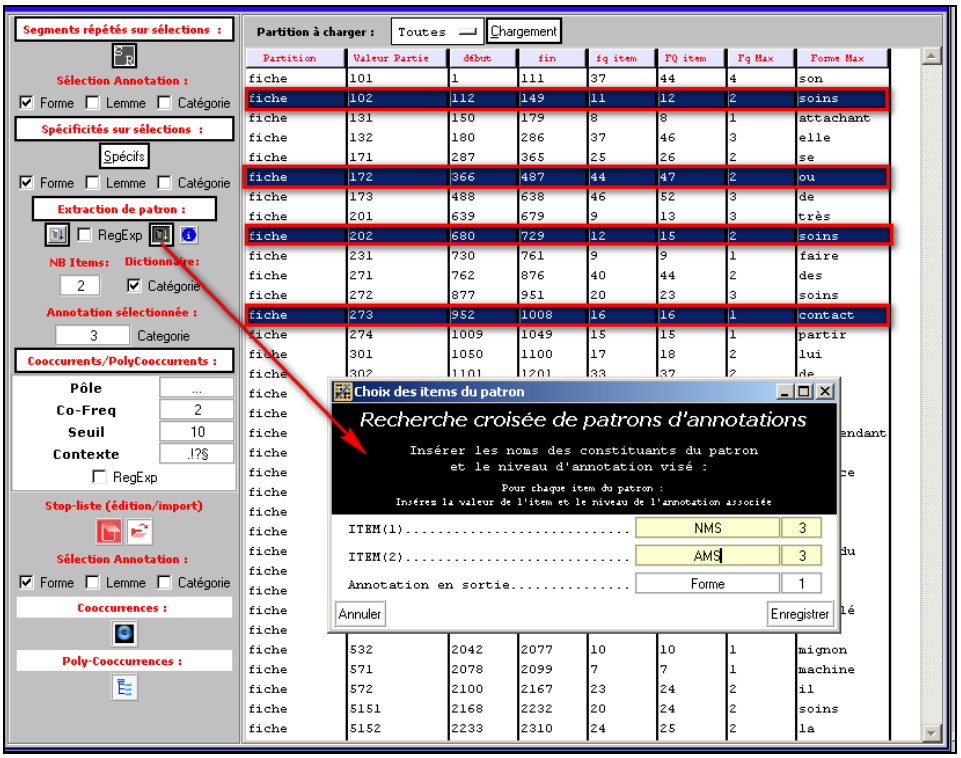
Specificites du vocabulaire - forme - sur selection de parties : fiche_102 fiche_172 fiche_202 fiche_273

Nombre d'occurrences du texte global = 2548
Nombre d'occurrences dans la partie visee = 90
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit :
C:/SERGEF~1/SF-WORK/07-DEVEL/ANR-TE~1/MKGRAP~1/BASES-~1/BASE-P~1/export/resultspecif-source-1334662323.txt

Forme	Ind-Specif	Fq-Totale	Fq-Partie
j'	2.5	1	1
ai	2.5	1	1
teint	2.5	1	1
pâle	2.5	1	1
situation	2.5	1	1
souriant	2.5	1	1
forte	2.5	1	1
jusqu'	2.5	1	1
lache	2.5	1	1
impossibilité	2.5	1	1
mis	2.5	1	1
mauvais	2.5	1	1
interaction	2.5	1	1
couverture	2.5	1	1
bain	2.5	1	1
mâchonne	2.5	1	1
sensible	2.5	1	1
accrocheur	2.5	1	1
créer	2.5	1	1
timide	2.5	1	1
exigeante	2.5	1	1
hypotonique	2.5	1	1
raide	2.5	1	1
douleur	2.5	1	1
convient	2.5	1	1
infecté	2.5	1	1
ou	2.4	9	2
suit	2.2	2	1
détendu	2.2	12	2
infirmière	2.2	2	1
oedématié	2.2	2	1

Figure 149 : *Le Trameur* calcul de spécificités sur une sélection de parties

21.5.3 Extraction de patron



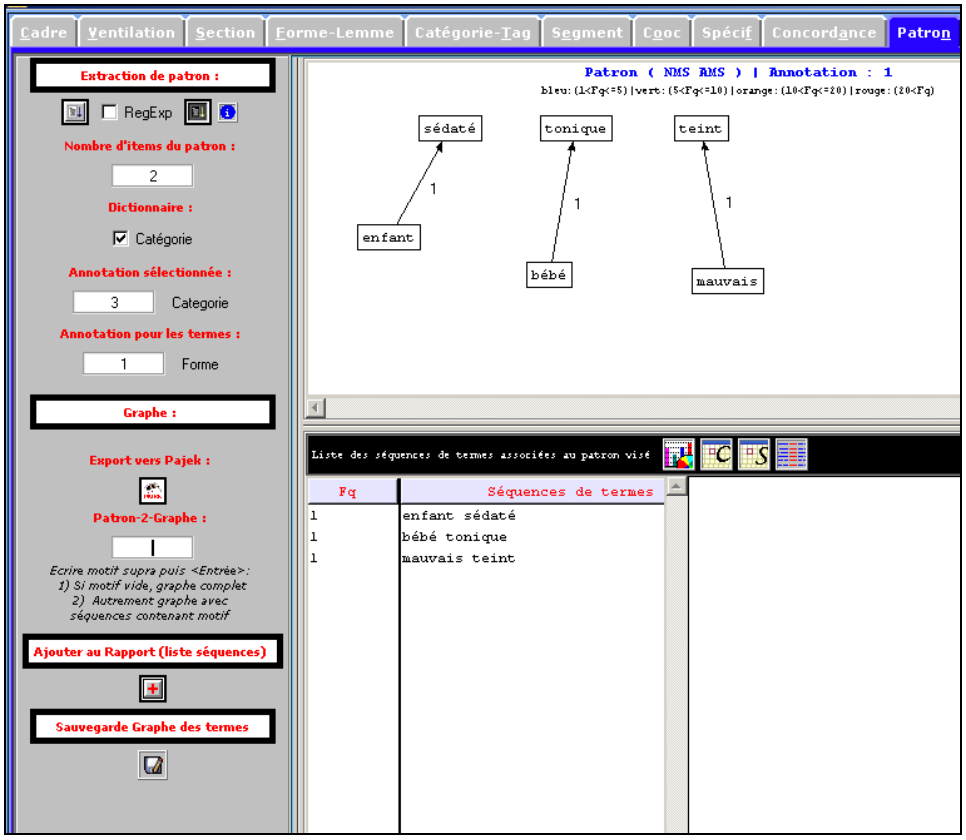


Figure 150 : *Le Trameur* extraction de patron sur une partie sélectionnée

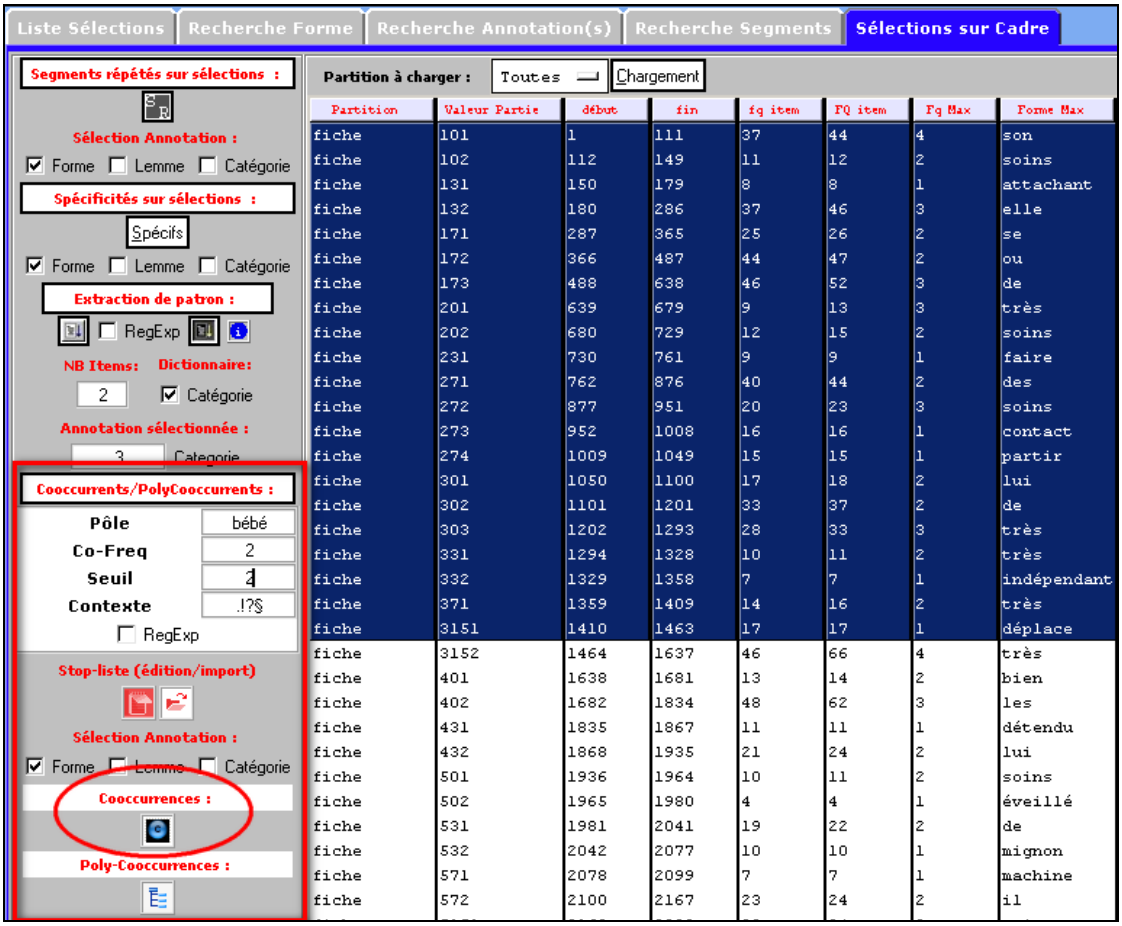
Comme précédemment, une concordance des termes produits illustrent la localisation de ces derniers sur la ***Trame*** (dans l'une des parties sélectionnées) :

```
------------------------------PARTIE{fiche=172}------------------------------
                    , attachant . bébé tonique et dynamique .
------------------------------PARTIE{fiche=202}------------------------------
                  hypotonique . aréactif . mauvais teint . labile .
------------------------------PARTIE{fiche=273}------------------------------
                          à voir . enfant sédaté , endormi .
```

Figure 151 : *Le Trameur* extraction de patron sur une partie sélectionnée

21.5.4 Calcul de cooccurrents

Cas n°1 : unité contextuelle déterminée par des délimiteurs


Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Sélections sur Cadre
Segments répétés sur sélections :
Sélection Annotation :
Forme Lemme Catégorie
Spécificités sur sélections :
Spécifs
Extraction de patron :
RegExp
NB Items:
Dictionnaire:
Catégorie
Annotation sélectionnée :
Cooccurrents/PolyCooccurrents :
Pôle bébé
Co-Freq 2
Seuil 2
Contexte .!?§
Stop-liste (édition/import)
Cooccurrences :
Poly-Cooccurrences :
Partition à charger : Toutes
Chargement
Partition | Valeur Partie | début | fin | fq item | FQ item | Fq Max | Forme Max
fiche | 101 | 1 | 111 | 37 | 44 | 4 | son
fiche | 102 | 112 | 149 | 11 | 12 | 2 | soins
fiche | 131 | 150 | 179 | 8 | 8 | 1 | attachant
fiche | 132 | 180 | 286 | 37 | 46 | 3 | elle
fiche | 171 | 287 | 365 | 25 | 26 | 2 | se
fiche | 172 | 366 | 487 | 44 | 47 | 2 | ou
fiche | 173 | 488 | 638 | 46 | 52 | 3 | de
fiche | 201 | 639 | 679 | 9 | 13 | 3 | très
fiche | 202 | 680 | 729 | 12 | 15 | 2 | soins
fiche | 231 | 730 | 761 | 9 | 9 | 1 | faire
fiche | 271 | 762 | 876 | 40 | 44 | 2 | des
fiche | 272 | 877 | 951 | 20 | 23 | 3 | soins
fiche | 273 | 952 | 1008 | 16 | 16 | 1 | contact
fiche | 274 | 1009 | 1049 | 15 | 15 | 1 | partir
fiche | 301 | 1050 | 1100 | 17 | 18 | 2 | lui
fiche | 302 | 1101 | 1201 | 33 | 37 | 2 | de
fiche | 303 | 1202 | 1293 | 28 | 33 | 3 | très
fiche | 331 | 1294 | 1328 | 10 | 11 | 2 | très
fiche | 332 | 1329 | 1358 | 7 | 7 | 1 | indépendant
fiche | 371 | 1359 | 1409 | 14 | 16 | 2 | très
fiche | 3151 | 1410 | 1463 | 17 | 17 | 1 | déplace
fiche | 3152 | 1464 | 1637 | 46 | 66 | 4 | très
fiche | 401 | 1638 | 1681 | 13 | 14 | 2 | bien
fiche | 402 | 1682 | 1834 | 48 | 62 | 3 | les
fiche | 431 | 1835 | 1867 | 11 | 11 | 1 | détendu
fiche | 432 | 1868 | 1935 | 21 | 24 | 2 | lui
fiche | 501 | 1936 | 1964 | 10 | 11 | 2 | soins
fiche | 502 | 1965 | 1980 | 4 | 4 | 1 | éveillé
fiche | 531 | 1981 | 2041 | 19 | 22 | 2 | de
fiche | 532 | 2042 | 2077 | 10 | 10 | 1 | mignon
fiche | 571 | 2078 | 2099 | 7 | 7 | 1 | machine
fiche | 572 | 2100 | 2167 | 23 | 24 | 2 | il

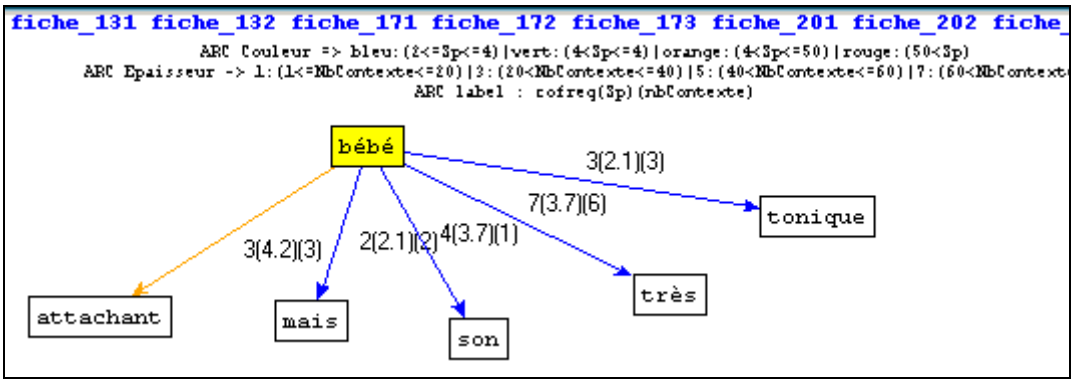


Figure 152 : *Le Trameur* cooccurrents/polycooccurrents sur une partie sélectionnée (1)

Le calcul des polycooccurrents peut ensuite être lancé.

Cas n°2 : unité contextuelle déterminée par le ***Cadre***

On reprend ici l'exemple de texte de travail vu plus haut :

```xml
<?xml version="1.0" encoding="iso-8859-1"?>
<?xml-stylesheet type="text/xsl" href="q2.xsl"?>
<TEI>
    <teiHeader>
        <fileDesc>
            <titleStmt>
                <title>Poésies</title>
                <author>Arthur Rimbaud</author>
            </titleStmt>
        </fileDesc>
    </teiHeader>
    <text>
        <group>
            <text>
                <front>
                    <head>Le Dormeur du Val</head>
                    <dateline>7 octobre 1870</dateline>
                </front>
                <body>
                    <lg>
                        <lg>
                            <l>C'est un trou de verdure où chante une rivière</l>
                            <l>Accrochant follement aux herbes des haillons</l>
                            <l>D'argent ; où le soleil, de la montagne fière,</l>
                            <l>Luit : c'est un petit val qui mousse de rayons.</l>
                        </lg>
                        <lg>
                            <l>Un soldat jeune, bouche ouverte, tête nue,</l>
                            <l>Et la nuque baignant dans le frais cresson bleu,</l>
                            <l>Dort ; il est étendu dans l'herbe, sous la nue,</l>
                            <l>Pâle dans son lit vert où la lumière pleut.</l>
                        </lg>
                        <lg>
                            <l>Les pieds dans les glaïeuls, il dort. Souriant comme</l>
                            <l>Sourirait un enfant malade, il fait un somme :</l>
                            <l>Nature, berce-le chaudement : il a froid.</l>
                        </lg>
                        <lg>
                            <l>Les parfums ne font pas frissonner sa narine ;</l>
                            <l>Il dort dans le soleil, la main sur sa poitrine</l>
                            <l>Tranquille. Il a deux trous rouges au côté droit.</l>
                        </lg>
                    </lg>
                </body>
            </text>
        </group>
    </text>
</TEI>
```

Figure 153 : *Le Trameur* cooccurrents/polycooccurrents sur une partie sélectionnée (2)

Dans la figure qui suit, la partition `lg` a été chargée dans le *Gestionnaire de Sélection* et certaines instances de strophes sélectionnées ; seules les strophes sélectionnées seront utilisées pour définir les contextes utilisés par le calcul.

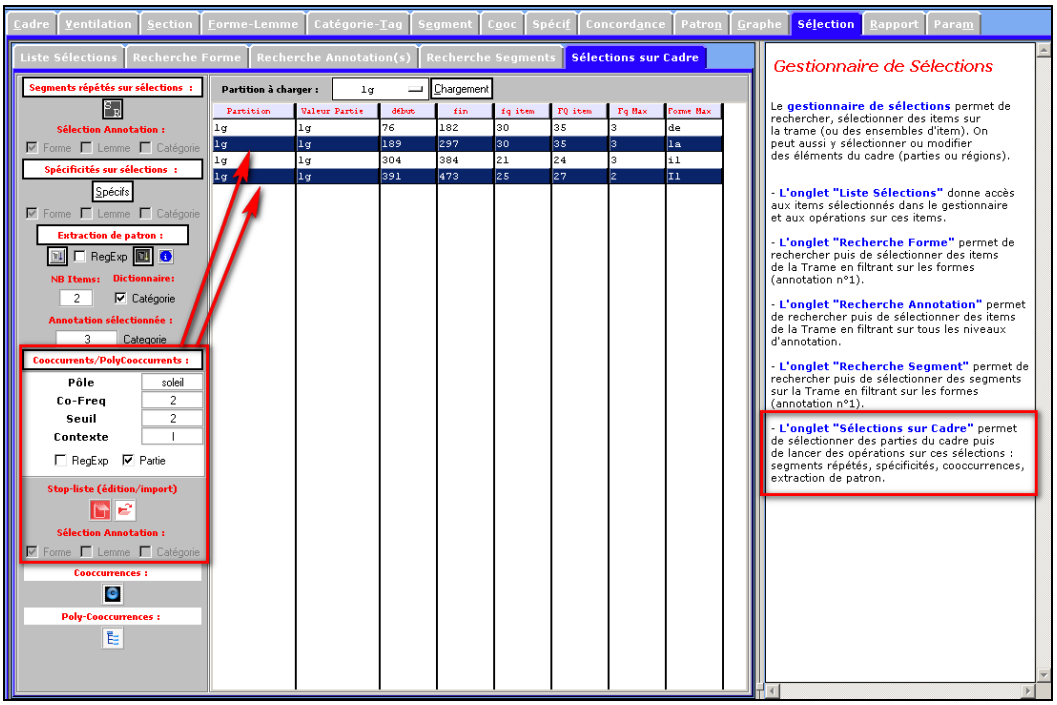

Cadre
Ventilation
Section
Forme-Lemme
Catégorie-Tag
Segment
Cooc
Spécif
Concordance
Patron
Graphe
Sélection
Rapport
Param
Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Sélections sur Cadre
Segments répétés sur sélections :
Sélection Annotation :
Spécificités sur sélections :
Spécifs
Extraction de patron :
RegExp
NB Items:
Dictionnaire:
Catégorie
Annotation sélectionnée :
Cooccurrents/PolyCooccurrents :
Pôle
soleil
Co-Freq
Seuil
Contexte
RegExp
Partie
Stop-liste (édition/import)
Cooccurrences :
Poly-Cooccurrences :
Partition à charger :
Chargement
Gestionnaire de Sélections
Le gestionnaire de sélections permet de rechercher, sélectionner des items sur la trame (ou des ensembles d'item). On peut aussi y sélectionner ou modifier des éléments du cadre (parties ou régions).
- L'onglet "Liste Sélections" donne accès aux items sélectionnés dans le gestionnaire et aux opérations sur ces items.
- L'onglet "Recherche Forme" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur les formes (annotation n°1).
- L'onglet "Recherche Annotation" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur tous les niveaux d'annotation.
- L'onglet "Recherche Segment" permet de rechercher puis de sélectionner des segments sur la Trame en filtrant sur les formes (annotation n°1).
- L'onglet "Sélections sur Cadre" permet de sélectionner des parties du cadre puis de lancer des opérations sur ces sélections : segments répétés, spécificités, cooccurrences, extraction de patron.

Figure 154 : *Le Trameur* cooccurrents/polycooccurrents sur une partie sélectionnée (3)

Comme précédemment, le calcul des cooccurrents utilisent des paramètres similaires. On peut soit définir des unités contextuelles (au sein des parties sélectionnées) en mettant en avant des caractères délimiteurs (*cf* cas n°1 *supra*), soit choisir d'utiliser une zone du ***Cadre*** pour déterminer les contextes.

Si cette seconde option est choisie, il est important de s'assurer que la partie utilisée pour définir le contexte est une sous-partie des parties sélectionnées dans le ***Cadre***.

Dans la figure précédente, les parties du ***Cadre*** sélectionnées correspondent à la première et à la troisième strophe ; la partie utilisée pour définir les contextes est la balise `l`, les contextes retenus correspondent donc aux vers constitutifs de ces 2 strophes :

```
Un soldat jeune, bouche ouverte, tête nue,
Et la nuque baignant dans le frais cresson bleu,
Dort ; il est étendu dans l'herbe, sous la nue,
Pâle dans son lit vert où la lumière pleut.
Les pieds dans les glaïeuls, il dort. Souriant comme
Sourirait un enfant malade, il fait un somme :
Nature, berce-le chaudement : il a froid.
```

Et au final, les contextes retenus pour le calcul rassemblent les vers contenant le pôle : dans le cas présent, aucun !

21.5.5 Calcul de cooccurrents sur corpus chronologique

On considère dans cette partie que l'on travaille sur le corpus chronologique « Le Père Duchesne » (*cf* Documentation en ligne Lexico3[21]), noté ici DUCHN.

[21] http://www.tal.univ-paris3.fr/lexico/manuels.htm

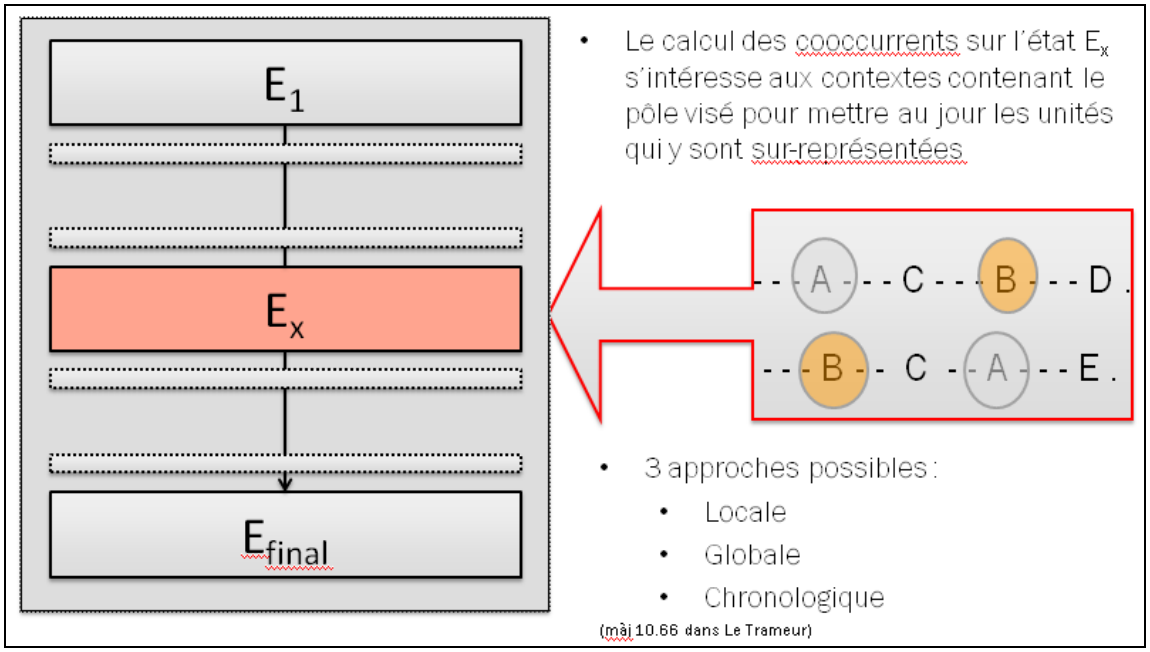

E1
Ex
Efinal
Le calcul des cooccurrents sur l'état Ex s'intéresse aux contextes contenant le pôle visé pour mettre au jour les unités qui y sont sur-représentées
- - A - - C - - B - - D .
- - B - - C - - A - - E .
3 approches possibles :
Locale
Globale
Chronologique
(màj 10.66 dans Le Trameur)

Figure 155 : Cooccurrents, approche locale / globale / chronologique

Approche « locale » : on travaille sur une sélection de partie pour une partition donnée

Partition sélectionnée : MOIS
Sélection de parties déclinées ci-dessous.
Calcul des cooccurrents de « patriotes »

Cas n°1 : 1 partie sélectionnée (MOIS=6), cooccurrence « locale » sur cette partie

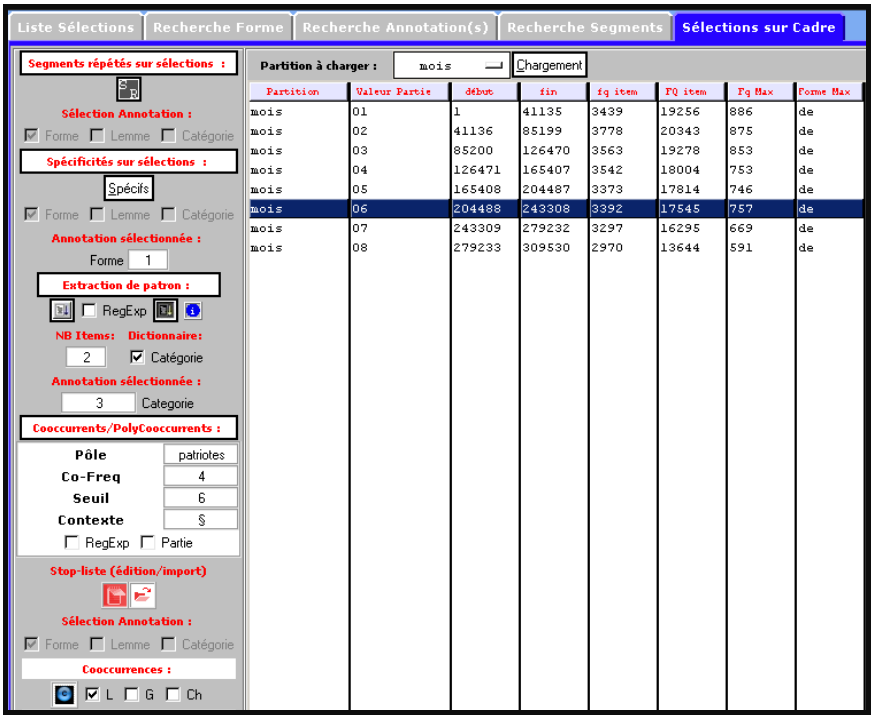

Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Sélections sur Cadre
Segments répétés sur sélections :
Sélection Annotation :
Forme Lemme Catégorie
Spécificités sur sélections :
Spécifs
Annotation sélectionnée :
Forme 1
Extraction de patron :
RegExp
NB Items: Dictionnaire:
2 Catégorie
Annotation sélectionnée :
3 Categorie
Cooccurrents/PolyCooccurrents :
Pôle patriotes
Co-Freq 4
Seuil 6
Contexte §
RegExp Partie
Stop-liste (édition/import)
Sélection Annotation :
Forme Lemme Catégorie
Cooccurrences :
L G Ch
Partition à charger : mois
Chargement
Partition | Valeur Partie | début | fin | fq item | FQ item | Fq Max | Forme Max
mois | 01 | 1 | 41135 | 3439 | 19256 | 886 | de
mois | 02 | 41136 | 85199 | 3778 | 20343 | 875 | de
mois | 03 | 85200 | 126470 | 3563 | 19278 | 853 | de
mois | 04 | 126471 | 165407 | 3542 | 18004 | 753 | de
mois | 05 | 165408 | 204487 | 3373 | 17814 | 746 | de
mois | 06 | 204488 | 243308 | 3392 | 17545 | 757 | de
mois | 07 | 243309 | 279232 | 3297 | 16295 | 669 | de
mois | 08 | 279233 | 309530 | 2970 | 13644 | 591 | de

Figure 156 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (1)

- Option L (local) (activée en bas à gauche)
- On « oublie » le reste du corpus (pas de fond) : les paramètres lexicométriques globaux choisis sont ceux de cette partie.
- T : Nb occurrence de la partie
- t : nb occcurrence des contextes contenant le pôle
- F : fréquence d'une forme sur la partie
- f : fréquence d'une forme dans les contextes contenant le pôle

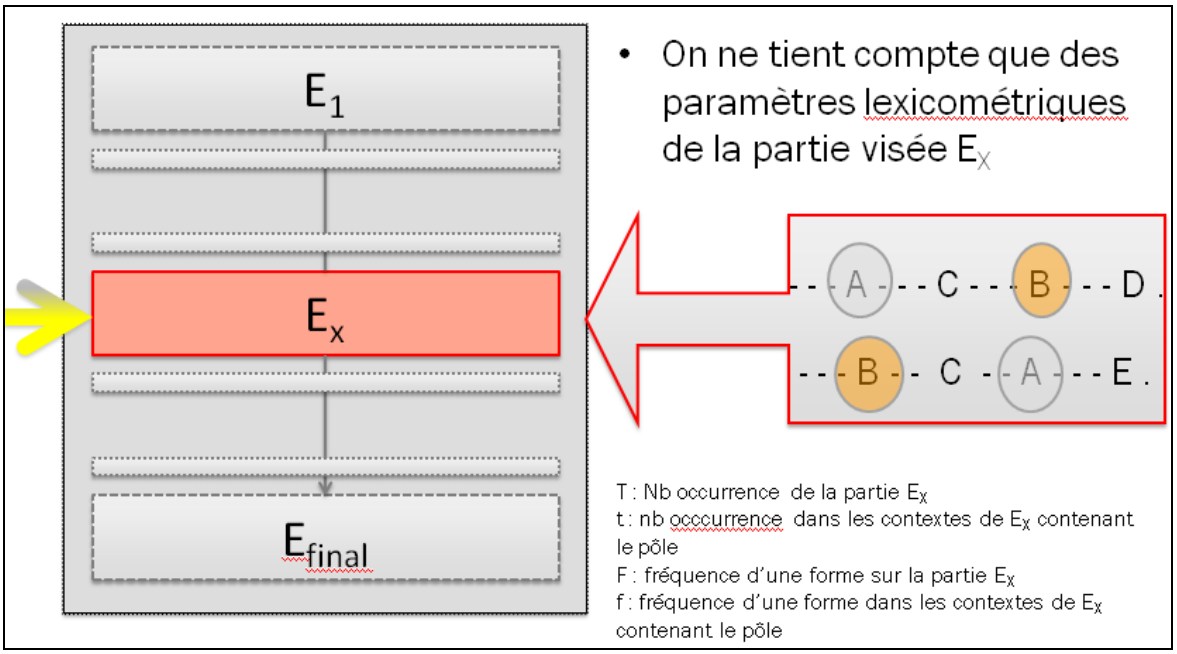

E1
Ex
Efinal
On ne tient compte que des paramètres lexicométriques de la partie visée EX
A C B D
B C A E
T : Nb occurrence de la partie EX
t : nb occcurrence dans les contextes de EX contenant le pôle
F : fréquence d'une forme sur la partie EX
f : fréquence d'une forme dans les contextes de EX contenant le pôle

Figure 157 : Cooccurrents, approche locale

Le réseau produit a l'allure suivante :

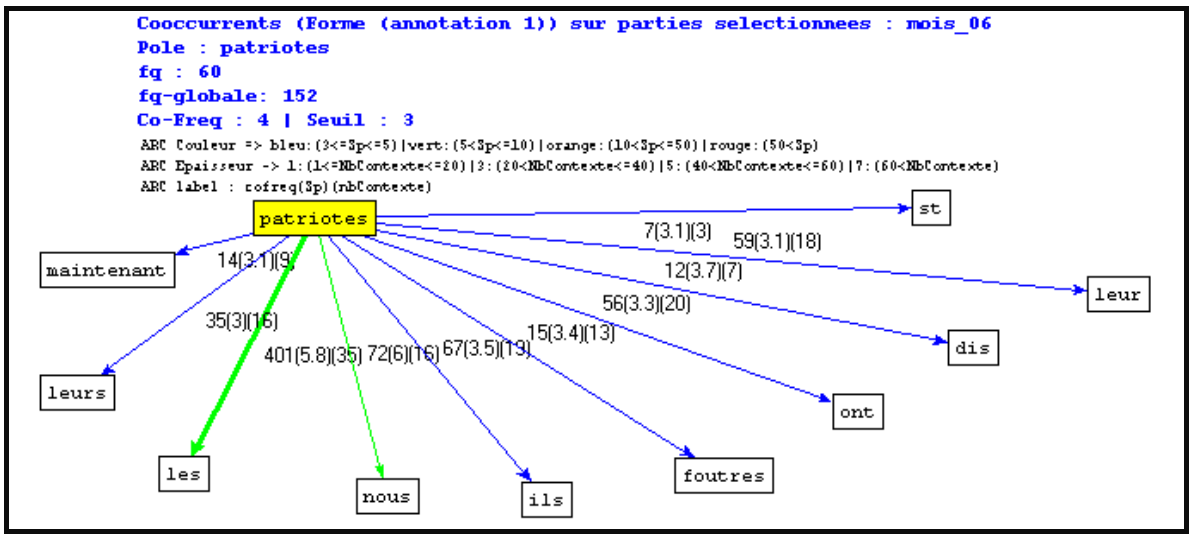


Figure 158 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (2)

Réitérer le même calcul sur d'autres parties permet d'obtenir un « réseau chronologique de cooccurrents » du même pôle.

Cas n°2 : 1 partie sélectionnée (MOIS=6), cooccurrence « globale » sur cette partie

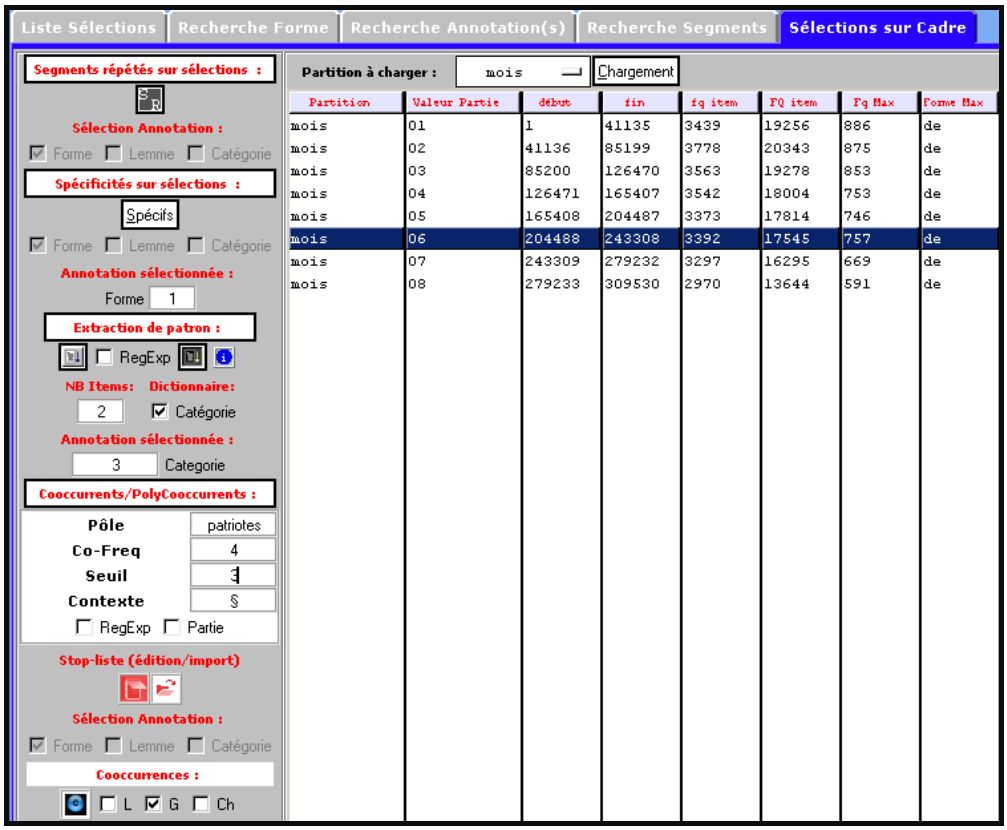


Figure 159 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (3)

- Option G (global) (activée en bas à gauche)
- On conserve le reste du corpus (prise en compte du fond) : les paramètres lexicométriques globaux choisis sont ceux du corpus complet.
- T : Nb occurrence du corpus complet
- t : nb occcurrence des contextes contenant le pôle
- F : fréquence d'une forme dans le corpus complet
- f : fréquence d'une forme dans les contextes contenant le pôle

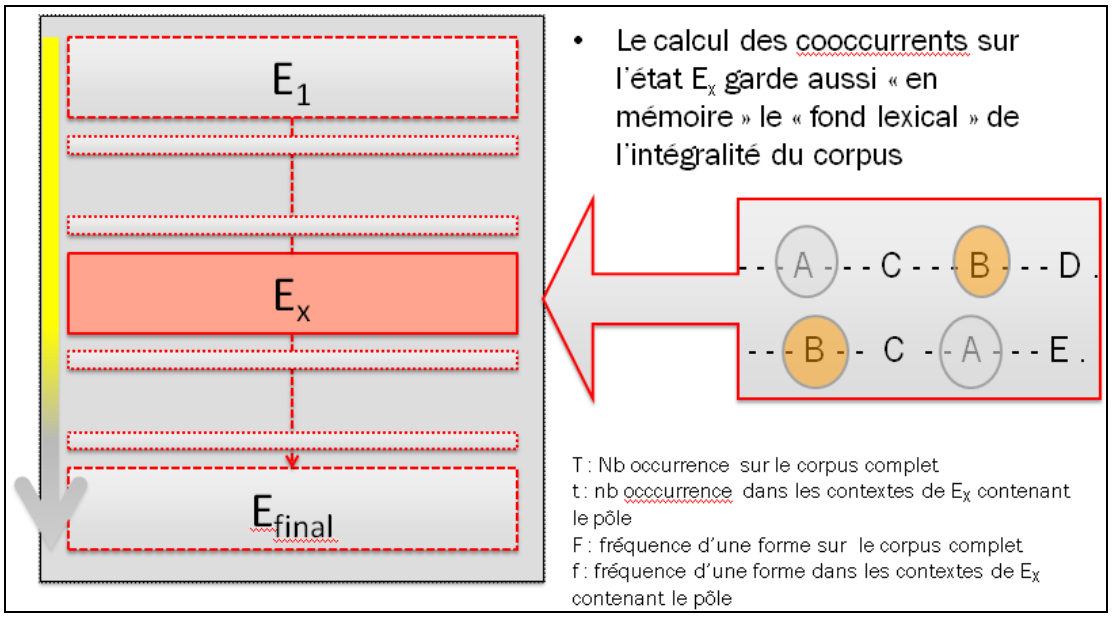

E1
Ex
Efinal
Le calcul des cooccurrents sur l'état Ex garde aussi « en mémoire » le « fond lexical » de l'intégralité du corpus
A C B D .
B C A E .
T : Nb occurrence sur le corpus complet
t : nb occcurrence dans les contextes de Ex contenant le pôle
F : fréquence d'une forme sur le corpus complet
f : fréquence d'une forme dans les contextes de Ex contenant le pôle

Figure 160 : Cooccurrents, approche globale

Le réseau produit a l'allure suivante :

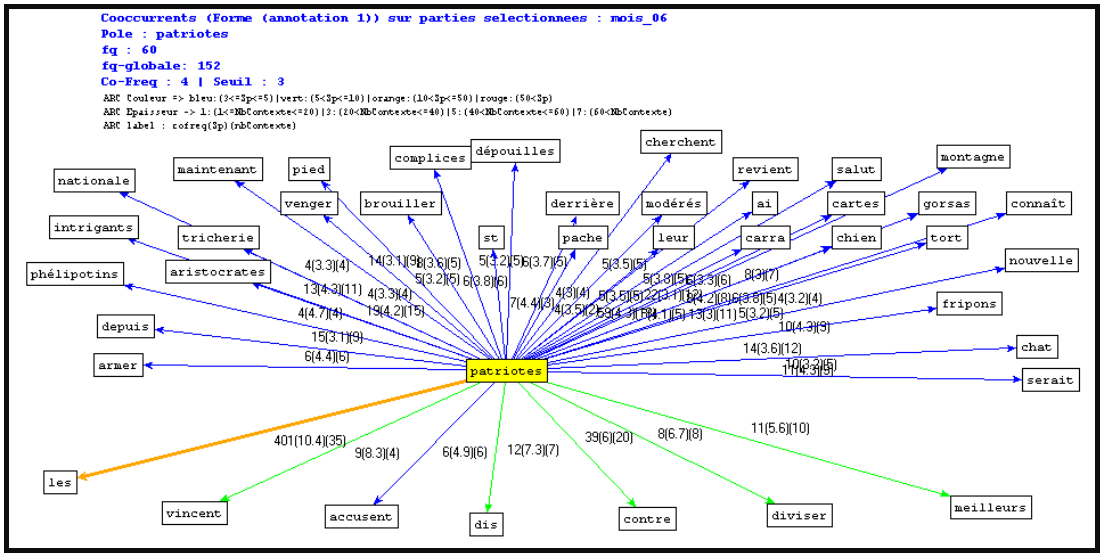
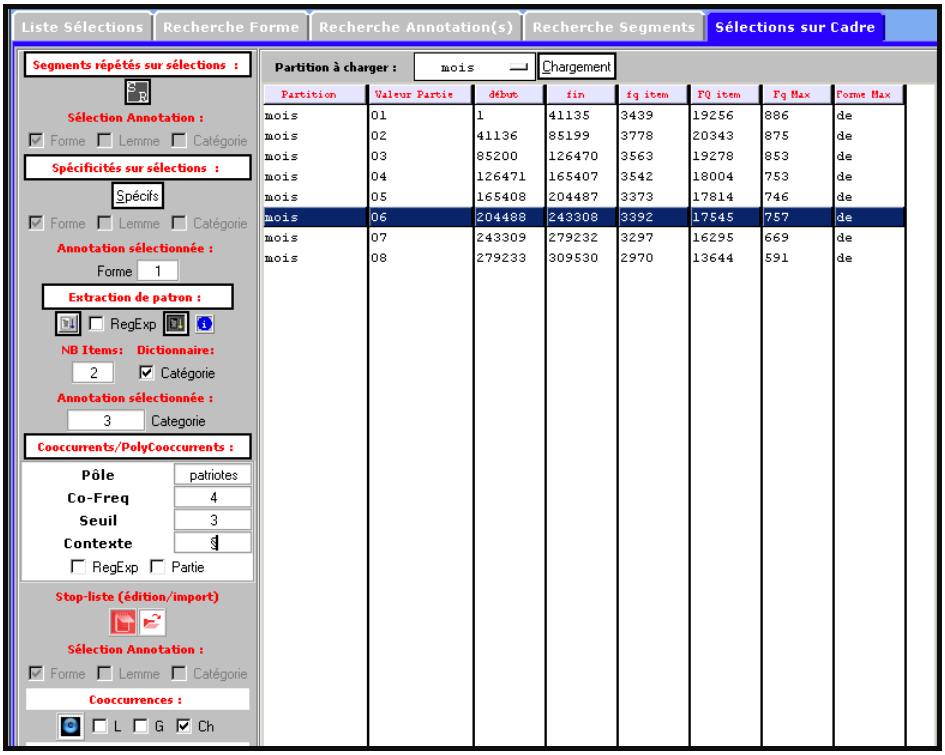


Figure 161 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (4)

Cas n°3 : 1 partie sélectionnée (MOIS=6), cooccurrence « chronologique » sur cette partie



Partition	Valeur Partie	début	fin	fq item	FQ item	Fq Max	Forme Max
mois	01	1	41135	3439	19256	886	de
mois	02	41136	85199	3778	20343	875	de
mois	03	85200	126470	3563	19278	853	de
mois	04	126471	165407	3542	18004	753	de
mois	05	165408	204487	3373	17814	746	de
mois	06	204488	243308	3392	17545	757	de
mois	07	243309	279232	3297	16295	669	de
mois	08	279233	309530	2970	13644	591	de

Figure 162 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (5)

- Option Ch (chrono) (activée en bas à gauche)
- On conserve les parties du corpus antérieures chronologiquement à celle choisie (prise en compte partielle du fond) : les paramètres lexicométriques globaux choisis sont ceux du corpus constitué par les parties antérieures et la partie sélectionnée.
- T : Nb occurrence dans les parties antérieures et la partie sélectionnée
- t : nb occcurrence des contextes contenant le pôle
- F : fréquence d'une forme dans les parties antérieures et la partie sélectionnée
- f : fréquence d'une forme dans les contextes contenant le pôle

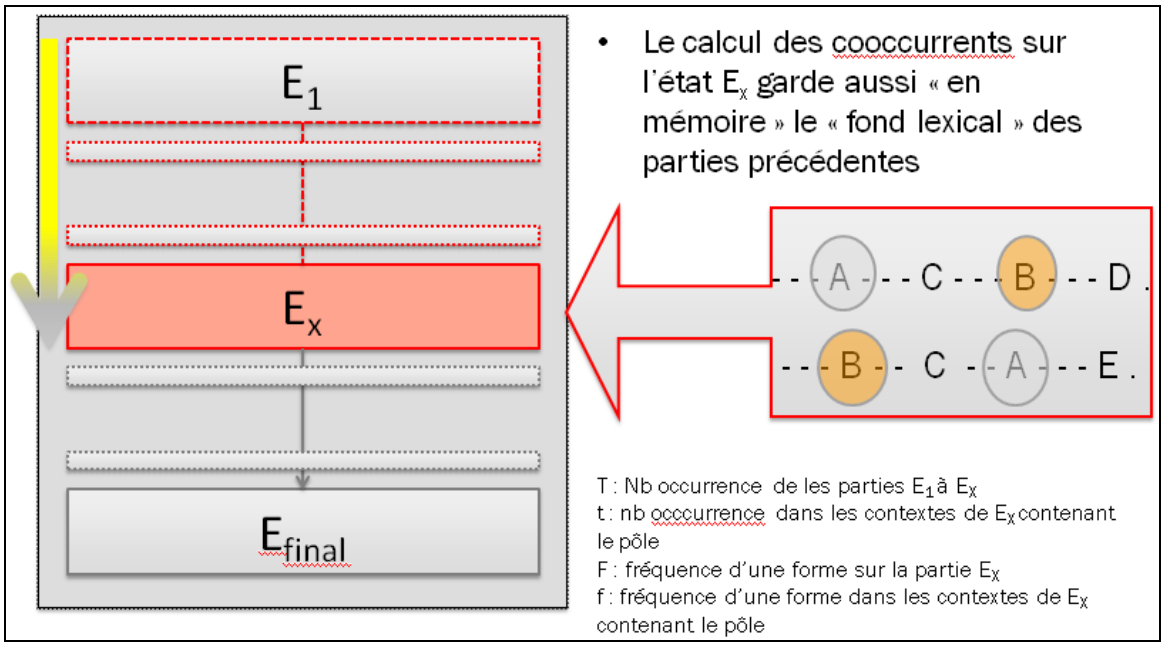

E1
Ex
Efinal
Le calcul des cooccurrents sur l'état Ex garde aussi « en mémoire » le « fond lexical » des parties précédentes
A C B D
B C A E
T : Nb occurrence de les parties E1 à Ex
t : nb occurrence dans les contextes de Ex contenant le pôle
F : fréquence d'une forme sur la partie Ex
f : fréquence d'une forme dans les contextes de Ex contenant le pôle

Figure 163 : Cooccurrents, approche chronologique

Le réseau produit a l'allure suivante :

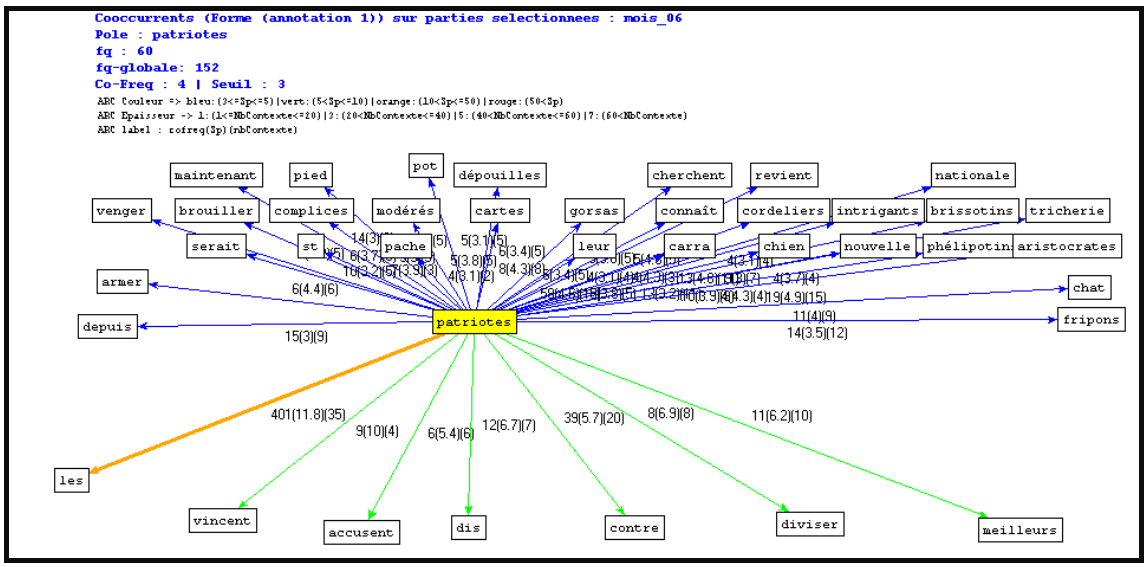


Figure 164 : *Le Trameur* cooccurrents sur une partie sélectionnée dans un corpus chronlogique (6)

La forme « accusent » serait ici un cooccurrent chronologiquement « très spécifique » de «patriotes » sur cette partie.

21.6 Gestion des sélections : sauvegarde et chargement

Un processus permet de **sauvegarder** des sélections : la sauvegarde d'une sélection permet de **nommer** cette sélection (ce nom étant ensuite utilisé pour activer ou inhiber les positions concernées par la sélection), de la **décrire** (édition d'une zone de texte pour qualifier de manière précise le contenu de la sélection).

Un autre processus permet aussi de **charger** une sélection préalablement construite. Le sélection chargée est ensuite *activable* ou *inhibable*.

Figure 165 : *Le Trameur* Gestion des sélections : sauvegarde et chargement

21.6.1 Sauvegarde d'une Sélection

Les figures qui suivent présentent le processus de sauvegarde.

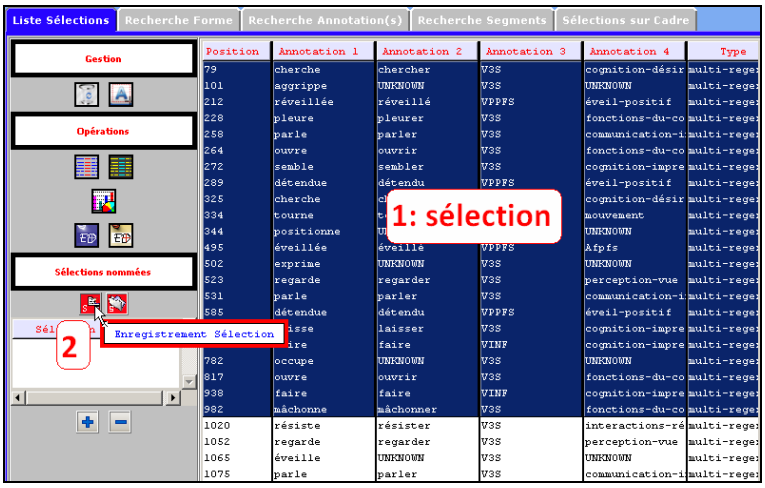


Figure 166 : *Le Trameur* sauvegarde d'une sélection (1)

Figure 167 : *Le Trameur* sauvegarde d'une sélection (2)

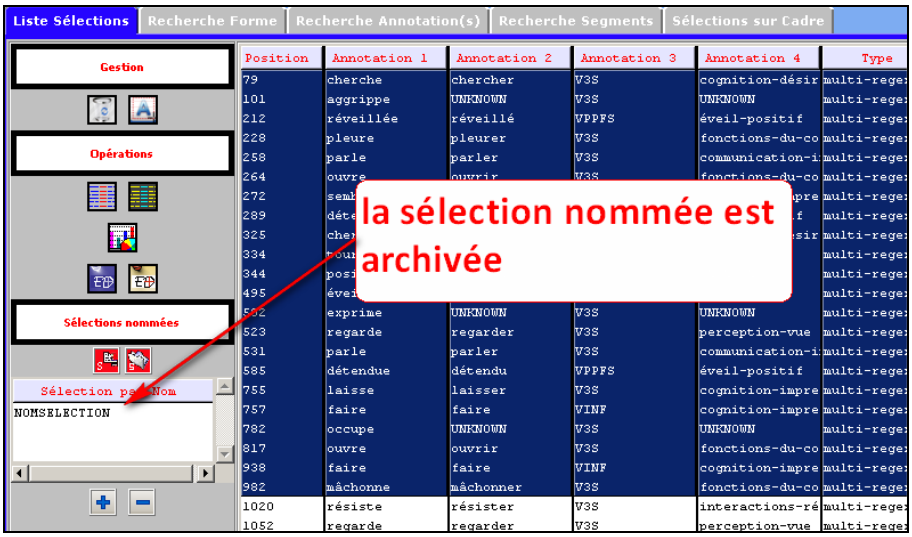


Figure 168 : *Le Trameur* sauvegarde d'une sélection (3)

Le processus décrit ci-dessus génère un fichier (dans le dossier `selections` du répertoire courant du ***Trameur***) regroupant les différentes informations identifiant la sélection.

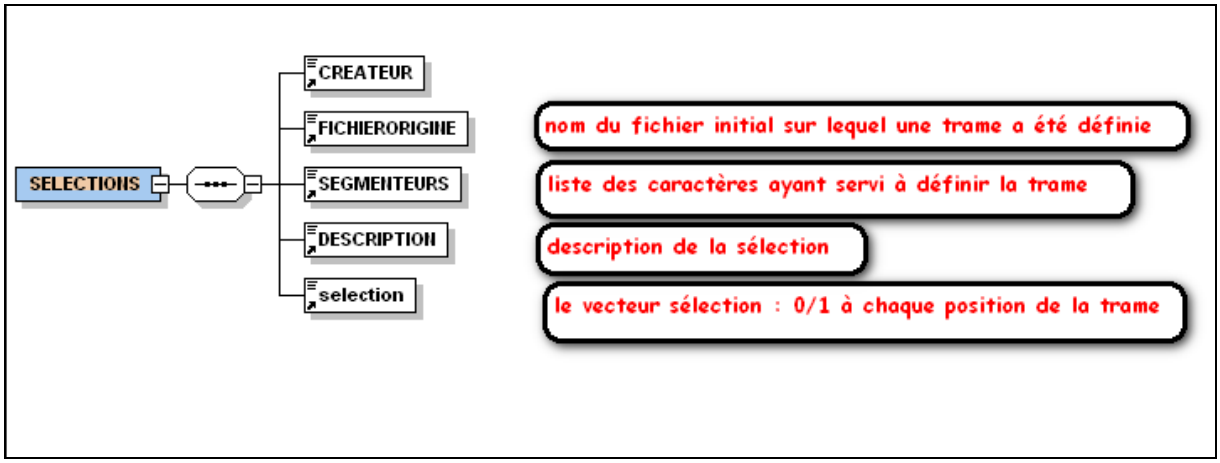
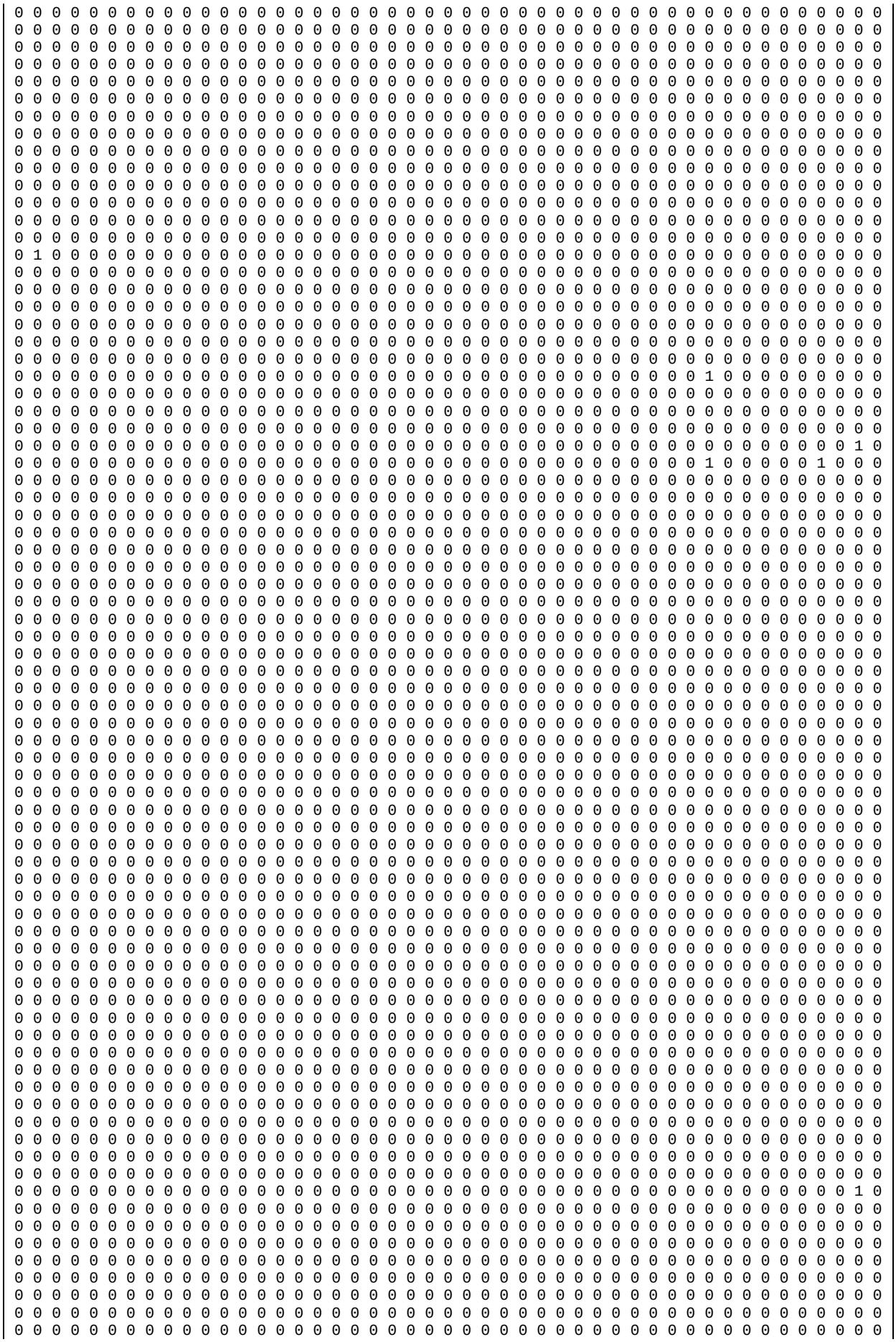


Figure 169 : *Le Trameur* arborescence XML d'une sélection

On donne ci-dessous le contenu d'un fichier produit (`selection_NOMSELECTION_1334663471.xml`), les positions sélectionnées ont ici été marquées « lisiblement » (1 rouge):

```
<?xml version="1.0" encoding="utf-8"?>
<SELECTIONS>
<CREATEUR>Le Trameur 8.00 (8.0b010)</CREATEUR>
<FICHIERORIGINE>base-trameur-p96-annot.txt</FICHIERORIGINE>
<SEGMENTEURS><![CDATA[ :,
?).(!]]></SEGMENTEURS>
<DESCRIPTION><![CDATA[la liste sélectionnée contient le résultat de la recherche de tous les
adjectifs associés Ã  l'annotation "émotion-bébé".
]]></DESCRIPTION>
<selection nom="emotion-bebe"  dimension="7181">
0 0 0 
```



```
0 0 0 
```

21.6.2 Chargement d'une Sélection

Dans les figures qui suivent, on charge une sélection (correspondant à celle enregistrée ci-dessus dans le fichier « `selection_NOMSELECTION_1334663471.xml` »).

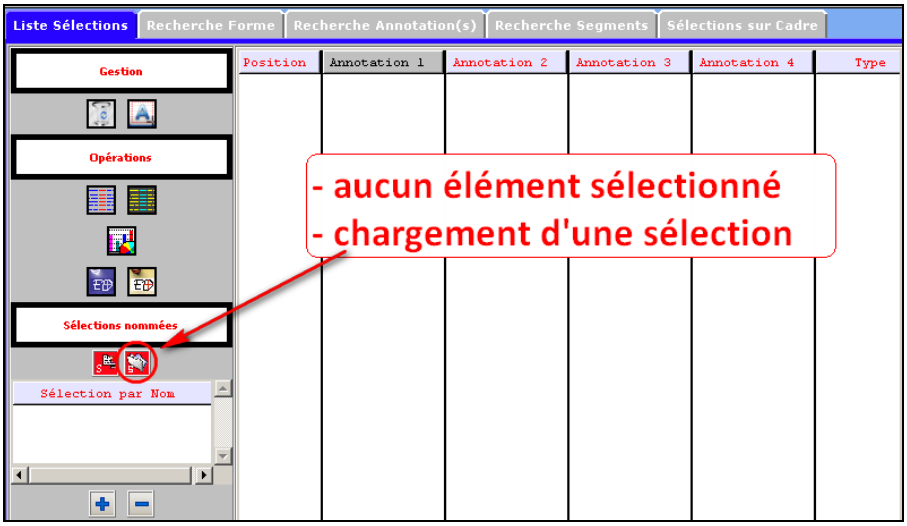


Figure 170 : *Le Trameur* chargement d'une sélection (1)

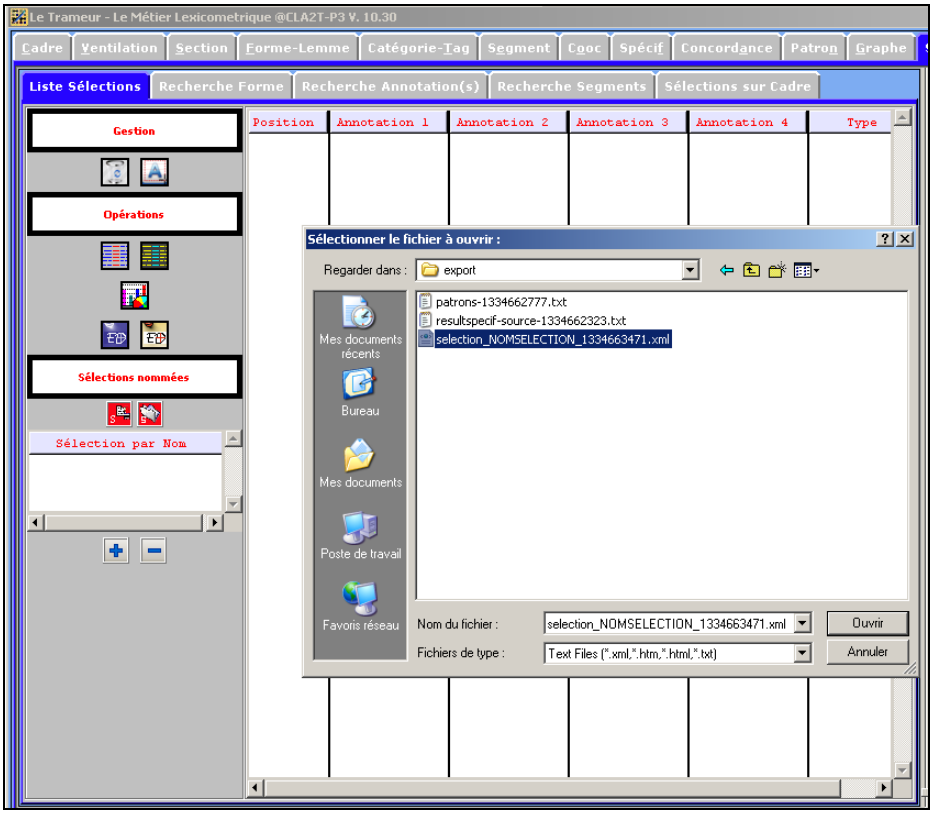


Figure 171 : *Le Trameur* chargement d'une sélection (2)

La sélection chargée peut ensuite être activée ou inhibée :

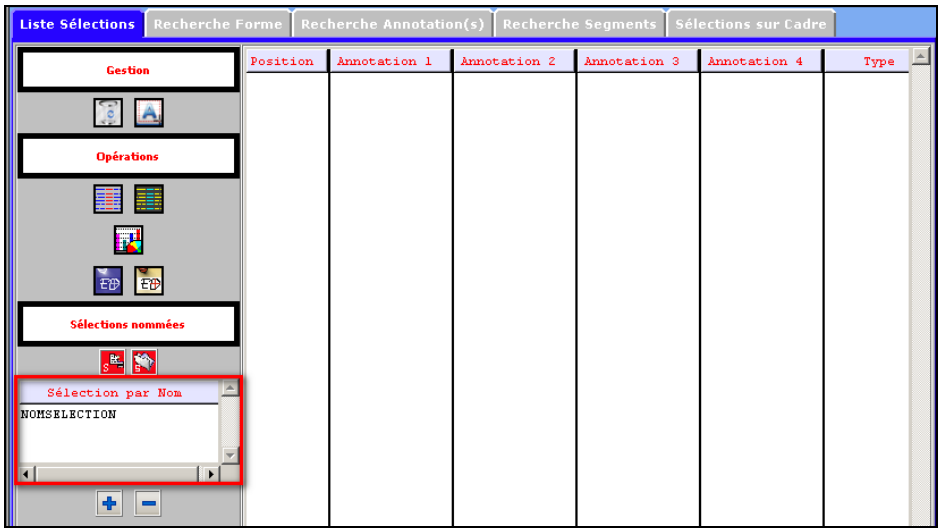


Figure 172 : *Le Trameur* activation d'une sélection nommée (1)

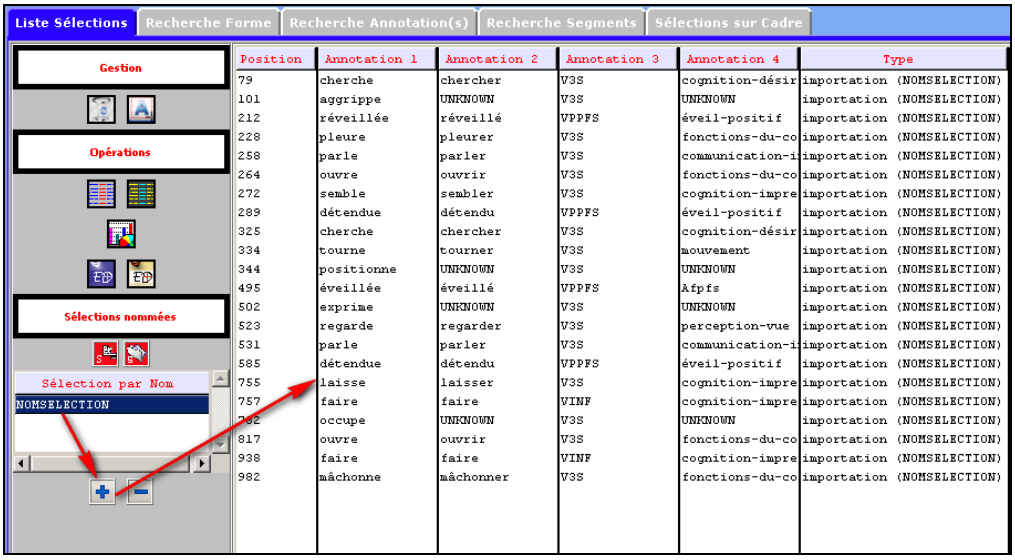

Liste Sélections
Recherche Forme
Recherche Annotation(s)
Recherche Segments
Sélections sur Cadre
Gestion
Opérations
Sélections nommées
Sélection par Nom
NOMSELECTION
Position Annotation 1 Annotation 2 Annotation 3 Annotation 4 Type
79 cherche chercher V3S cognition-désir importation (NOMSELECTION)
101 aggrippe UNKNOWN V3S UNKNOWN importation (NOMSELECTION)
212 réveillée réveillé VPPFS éveil-positif importation (NOMSELECTION)
228 pleure pleurer V3S fonctions-du-co importation (NOMSELECTION)
258 parle parler V3S communication-i importation (NOMSELECTION)
264 ouvre ouvrir V3S fonctions-du-co importation (NOMSELECTION)
272 semble sembler V3S cognition-impre importation (NOMSELECTION)
289 détendue détendu VPPFS éveil-positif importation (NOMSELECTION)
325 cherche chercher V3S cognition-désir importation (NOMSELECTION)
334 tourne tourner V3S mouvement importation (NOMSELECTION)
344 positionne UNKNOWN V3S UNKNOWN importation (NOMSELECTION)
495 éveillée éveillé VPPFS Afpfs importation (NOMSELECTION)
502 exprime UNKNOWN V3S UNKNOWN importation (NOMSELECTION)
523 regarde regarder V3S perception-vue importation (NOMSELECTION)
531 parle parler V3S communication-i importation (NOMSELECTION)
585 détendue détendu VPPFS éveil-positif importation (NOMSELECTION)
755 laisse laisser V3S cognition-impre importation (NOMSELECTION)
757 faire faire VINF cognition-impre importation (NOMSELECTION)
782 occupe UNKNOWN V3S UNKNOWN importation (NOMSELECTION)
817 ouvre ouvrir V3S fonctions-du-co importation (NOMSELECTION)
938 faire faire VINF cognition-impre importation (NOMSELECTION)
982 mâchonne mâchonner V3S fonctions-du-co importation (NOMSELECTION)

Figure 173 : *Le Trameur* activation d'une sélection nommée (2)

L'activation d'un *Clic-droit* sur la sélection nommée permet d'accéder à sa description (que l'on peut modifier ou enrichir).

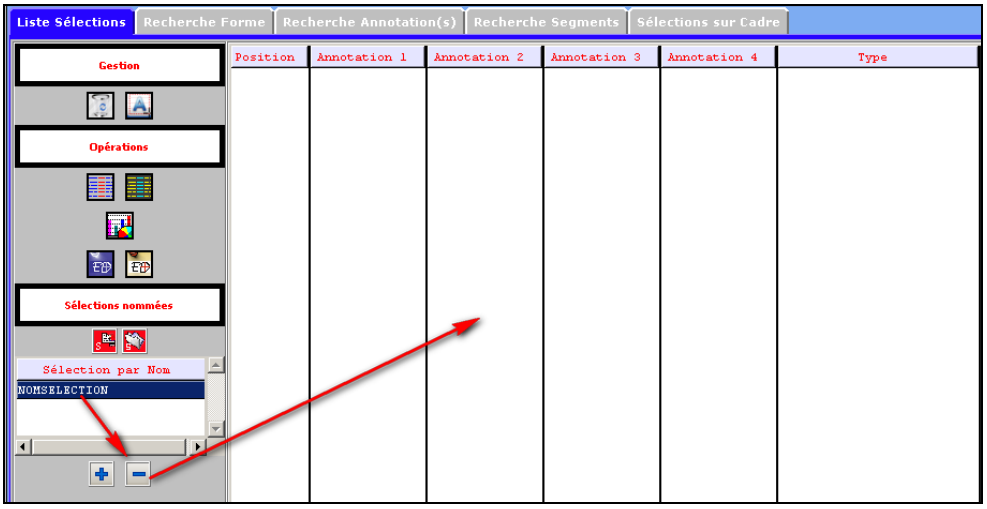


Figure 174 : *Le Trameur* inhibation d'une sélection nommée

22 Importation d'annotation sur une *Trame*

Il est possible d'importer un ou plusieurs systèmes d'annotations sur une ***Trame*** et un ***Cadre*** déjà chargés dans le ***Trameur***. Ce processus consiste à annoter chaque position de la ***Trame*** en lui associant une nouvelle « étiquette ».

Annotation importée n°1	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	A(7)	A(8)	
Catégorie		Cat(Le)		Cat(dormeur)		Cat(du)		Cat(val)	…
Lemme		Lemme(Le)		Lemme(dormeur)		Lemme(du)		Lemme(val)	…
Forme		**Le**		**dormeur**		**du**		**val**	…
Positions	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	…

Chaque item (défini par une position unique) est associé à sa forme graphique initiale (l'annotation n°1 dite « Forme »), éventuellement son lemme (l'annotation n°2 dite « Lemme ») et sa catégorie (l'annotation n°3 dite « Catégorie ») si *treetagger* a été paramétré, et une nouvelle « étiquette » issue du système d'annotations importé.

Le format du fichier d'annotations à importer s'appuie sur le fichier de tramage issu de l'export présenté *supra*.

Dans ce qui suit, nous considérons que nous allons associer une couleur à chaque item de la ***Trame*** :

couleur	bleu	vert	rouge	orange	bleu	rouge	jaune	bleu	
Forme		Le		dormeur		du		val	…
Positions	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	…

Le fichier complet d'import a l'allure suivante :

entête

```
<?xml version="1.0" encoding="iso-8859-1"?>
<importannotationsurtrame>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame), importable et
echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
<p>Samedi 26 Mai 2007
 12:02:30
  Ce document n'est pas encore publie.</p>
</publicationStmt>
<sourceDesc>
Fichier d'import d'annotations sur une Trame deja generee sur le fichier initial
Fichier          initialement          traite          :          C:/SFleury/DEVEL/WIP-ANR-
textometrrie/mkGraphTrameCadre/textes/dormeurL3.txt
Encodage : iso-8859-1
Delimiteur : \.\ \,\:\;\!\?\/\_\-\"\'\(\)\[\]\{\}\§\$\|\&amp;gt;\&amp;lt;\=\+

</sourceDesc>
</fileDesc>
</teiheader>
<items>
<item type="delim" pos="1"><a>bleu</a></item>
<item type="forme" pos="2"><a>vert</a></item>
<item type="delim" pos="3"><a>rouge</a></item>
<item type="forme" pos="4"><a>orange</a></item>
<item type="delim" pos="5"><a>bleu</a></item>
<item type="forme" pos="6"><a>rouge</a></item>
<item type="delim" pos="7"><a>jaune</a></item>
<item type="forme" pos="8"><a>bleu</a></item>
<item type="delim" pos="9"><a>noir</a></item>
<item type="delim" pos="10"><a>bleu</a></item>
<item type="delim" pos="11"><a>vert</a></item>
<item type="forme" pos="12"><a>vert</a></item>
<item type="delim" pos="13"><a>jaune</a></item>
<item type="delim" pos="14"><a>vert</a></item>
<item type="forme" pos="15"><a>rouge</a></item>
<item type="delim" pos="16"><a>bleu</a></item>
<item type="forme" pos="17"><a>vert</a></item>
<item type="delim" pos="18"><a>rouge</a></item>
<item type="forme" pos="19"><a>bleu</a></item>
<item type="delim" pos="20"><a>rouge</a></item>
<item type="forme" pos="21"><a>noir</a></item>
<item type="delim" pos="22"><a>jaune</a></item>
<item type="forme" pos="23"><a>noir</a></item>
<item type="delim" pos="24"><a>vert</a></item>
<item type="forme" pos="25"><a>bleu</a></item>
<item type="delim" pos="26"><a>rouge</a></item>
<item type="forme" pos="27"><a>noir</a></item>
<item type="delim" pos="28"><a>bleu</a></item>
<item type="forme" pos="29"><a>noir</a></item>
<item type="delim" pos="30"><a>noir</a></item>
<item type="forme" pos="31"><a>bleu</a></item>
<item type="delim" pos="32"><a>jaune</a></item>
<item type="delim" pos="33"><a>vert</a></item>
<item type="delim" pos="34"><a>rouge</a></item>
<item type="delim" pos="35"><a>jaune</a></item>
<item type="forme" pos="36"><a>bleu</a></item>
<item type="delim" pos="37"><a>bleu</a></item>
<item type="forme" pos="38"><a>vert</a></item>
<item type="delim" pos="39"><a>noir</a></item>
<item type="forme" pos="40"><a>rouge</a></item>
<item type="delim" pos="41"><a>bleu</a></item>
<item type="forme" pos="42"><a>noir</a></item>
<item type="delim" pos="43"><a>bleu</a></item>
<item type="forme" pos="44"><a>jaune</a></item>
<item type="delim" pos="45"><a>noir</a></item>
<item type="forme" pos="46"><a>bleu</a></item>
<item type="delim" pos="47"><a>bleu</a></item>
<item type="delim" pos="48"><a>vert</a></item>
```

Trame des annotations à importer

```
<item type="delim" pos="49"><a>bleu</a></item>
<item type="forme" pos="50"><a>rouge</a></item>
<item type="delim" pos="51"><a>noir</a></item>
<item type="delim" pos="52"><a>bleu</a></item>
<item type="forme" pos="53"><a>bleu</a></item>
<item type="delim" pos="54"><a>rouge</a></item>
<item type="delim" pos="55"><a>jaune</a></item>
<item type="delim" pos="56"><a>vert</a></item>
<item type="forme" pos="57"><a>bleu</a></item>
<item type="delim" pos="58"><a>rouge</a></item>
<item type="forme" pos="59"><a>bleu</a></item>
<item type="delim" pos="60"><a>noir</a></item>
<item type="forme" pos="61"><a>noir</a></item>
<item type="delim" pos="62"><a>bleu</a></item>
<item type="delim" pos="63"><a>noir</a></item>
<item type="delim" pos="64"><a>rouge</a></item>
<item type="forme" pos="65"><a>bleu</a></item>
<item type="delim" pos="66"><a>jaune</a></item>
<item type="forme" pos="67"><a>vert</a></item>
<item type="delim" pos="68"><a>rouge</a></item>
<item type="forme" pos="69"><a>jaune</a></item>
<item type="delim" pos="70"><a>vert</a></item>
<item type="forme" pos="71"><a>rouge</a></item>
<item type="delim" pos="72"><a>bleu</a></item>
<item type="delim" pos="73"><a>rouge</a></item>
<item type="delim" pos="74"><a>jaune</a></item>
<item type="delim" pos="75"><a>bleu</a></item>
<item type="forme" pos="76"><a>rouge</a></item>
<item type="delim" pos="77"><a>vert</a></item>
<item type="delim" pos="78"><a>jaune</a></item>
<item type="delim" pos="79"><a>noir</a></item>
<item type="forme" pos="80"><a>bleu</a></item>
<item type="delim" pos="81"><a>rouge</a></item>
<item type="delim" pos="82"><a>jaune</a></item>
<item type="forme" pos="83"><a>rouge</a></item>
<item type="delim" pos="84"><a>rouge</a></item>
<item type="forme" pos="85"><a>vert</a></item>
<item type="delim" pos="86"><a>bleu</a></item>
<item type="forme" pos="87"><a>jaune</a></item>
<item type="delim" pos="88"><a>rouge</a></item>
<item type="forme" pos="89"><a>bleu</a></item>
<item type="delim" pos="90"><a>vert</a></item>
<item type="forme" pos="91"><a>jaune</a></item>
<item type="delim" pos="92"><a>bleu</a></item>
<item type="forme" pos="93"><a>rouge</a></item>
<item type="delim" pos="94"><a>vert</a></item>
<item type="forme" pos="95"><a>bleu</a></item>
<item type="delim" pos="96"><a>jaune</a></item>
<item type="forme" pos="97"><a>rouge</a></item>
<item type="delim" pos="98"><a>noir</a></item>
<item type="delim" pos="99"><a>bleu</a></item>
<item type="forme" pos="100"><a>vert</a></item>
</items>
</importannotationsurtrame>
```

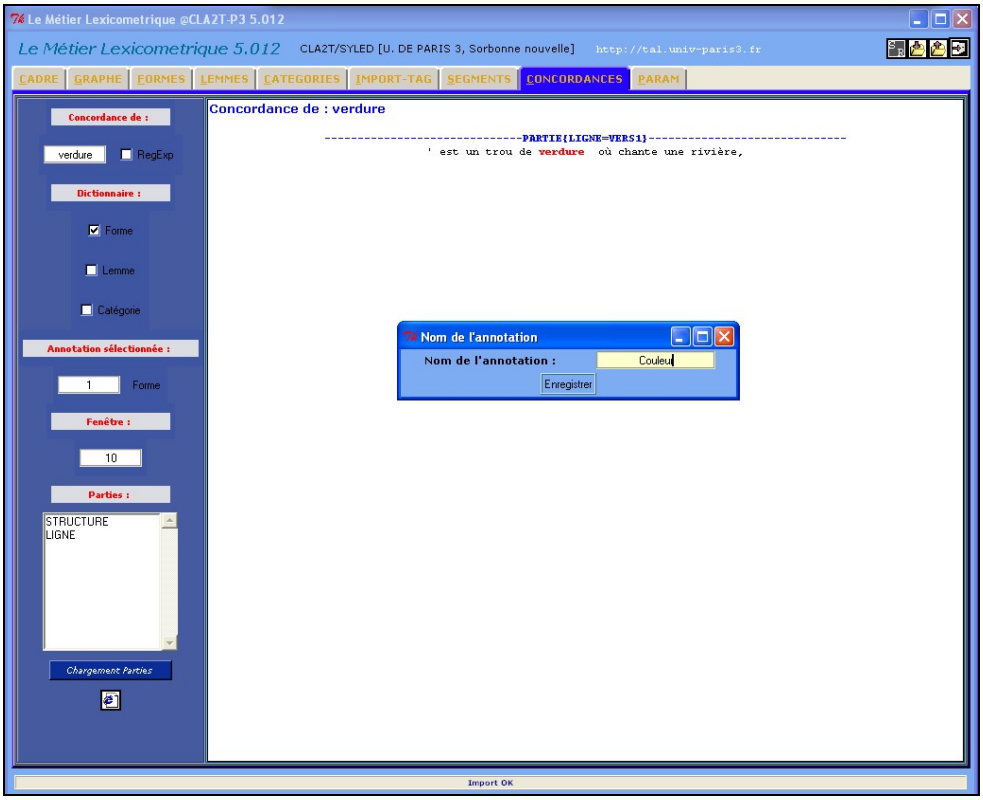
Pour charger ce fichier d'annotations, il suffit d'activer le bouton .

A la fin de l'importation il convient de nommer le système importé, ce nom sera affiché dans la zone « *Annotation sélectionnée* » dans les onglets **VENTILATION**, **CONCORDANCE** et **PATRON** quand le numéro de cette annotation sera sélectionné dans la zone de choix d'un numéro d'annotation.

Par convention :

- le numéro d'annotation n°1 correspond à « Forme »
- le numéro d'annotation n°2 correspond à « Lemme »
- le numéro d'annotation n°3 correspond à « Catégorie »

Toute annotation importée aura un numéro supérieur ou égal à 4.



(enregistrement du nom de l'annotation importée)

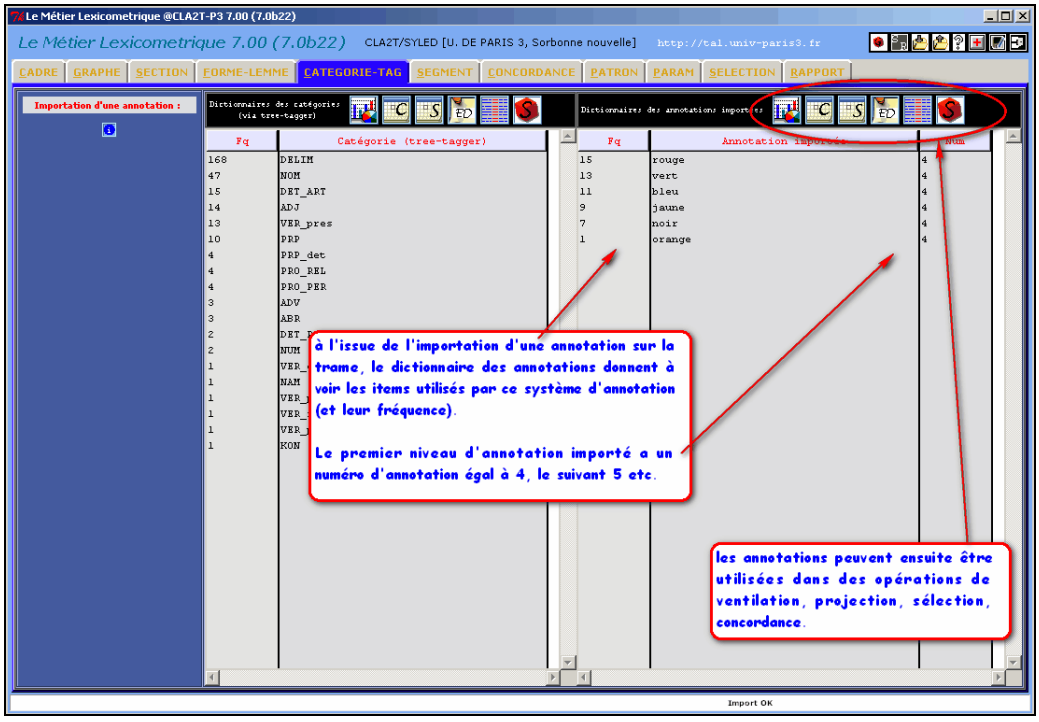


Figure 175 : ***Le Trameur*** **Importation d'annotation, nommage et sélection du niveau d'annotation**

Une fois chargé, ce nouveau système d'annotations est disponible pour les différentes opérations présentées précédemment : concordance, ventilation, recherche de patron, édition/modification des items.

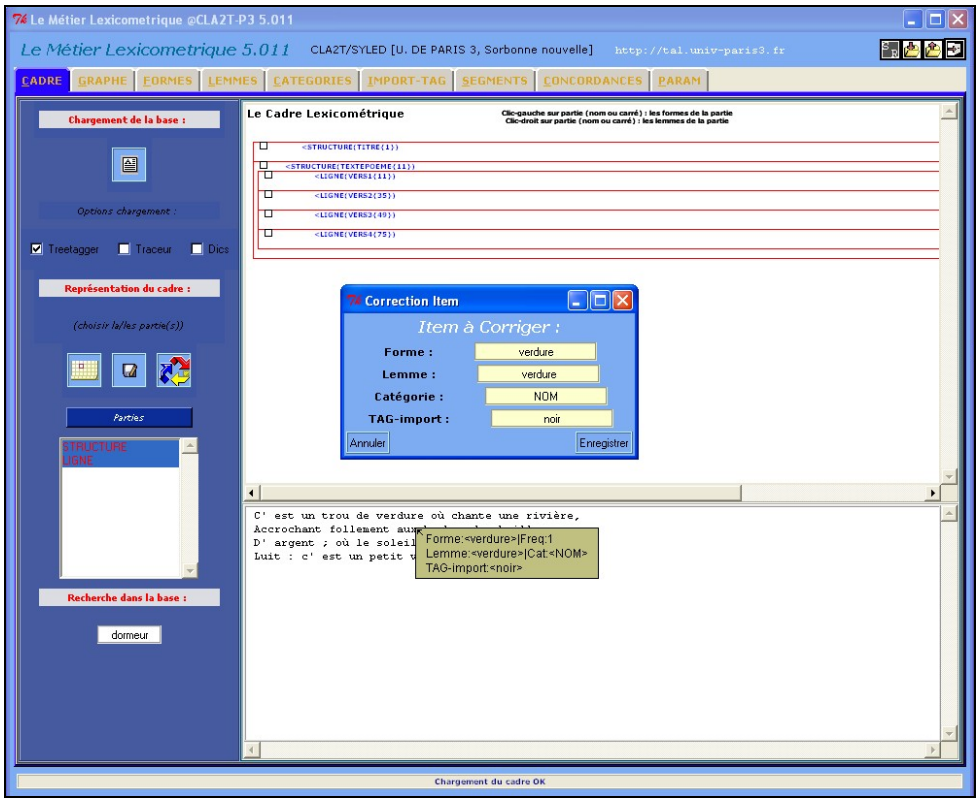


Figure 176 : *Le Trameur* correction d'item (annotation importée)

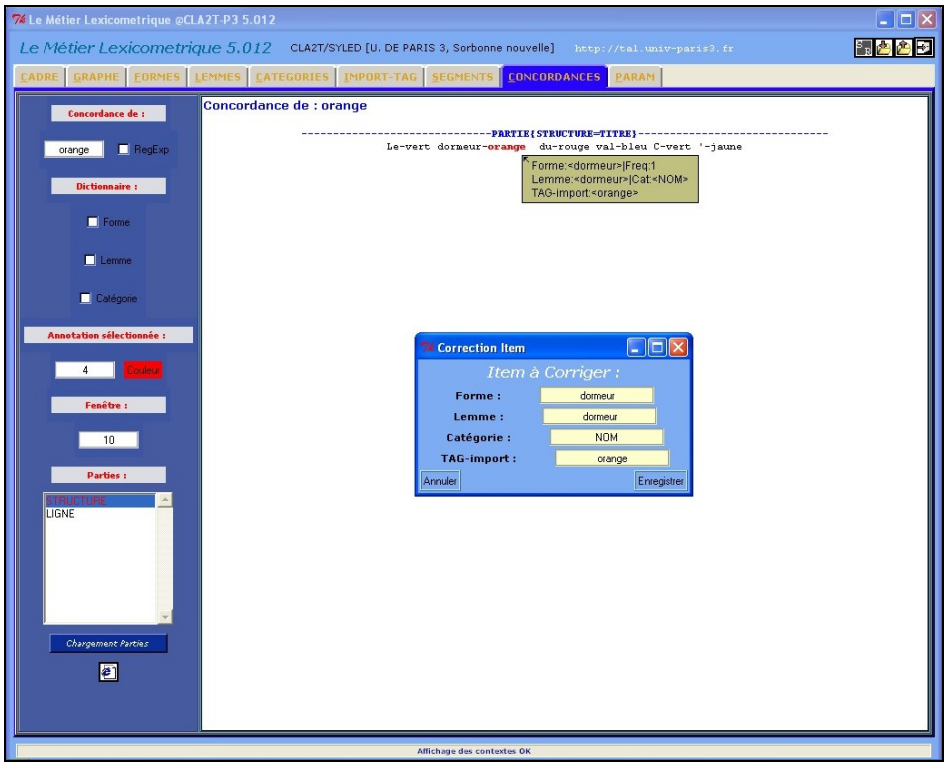


Figure 177 : *Le Trameur* concordance d'item (annotation importée)

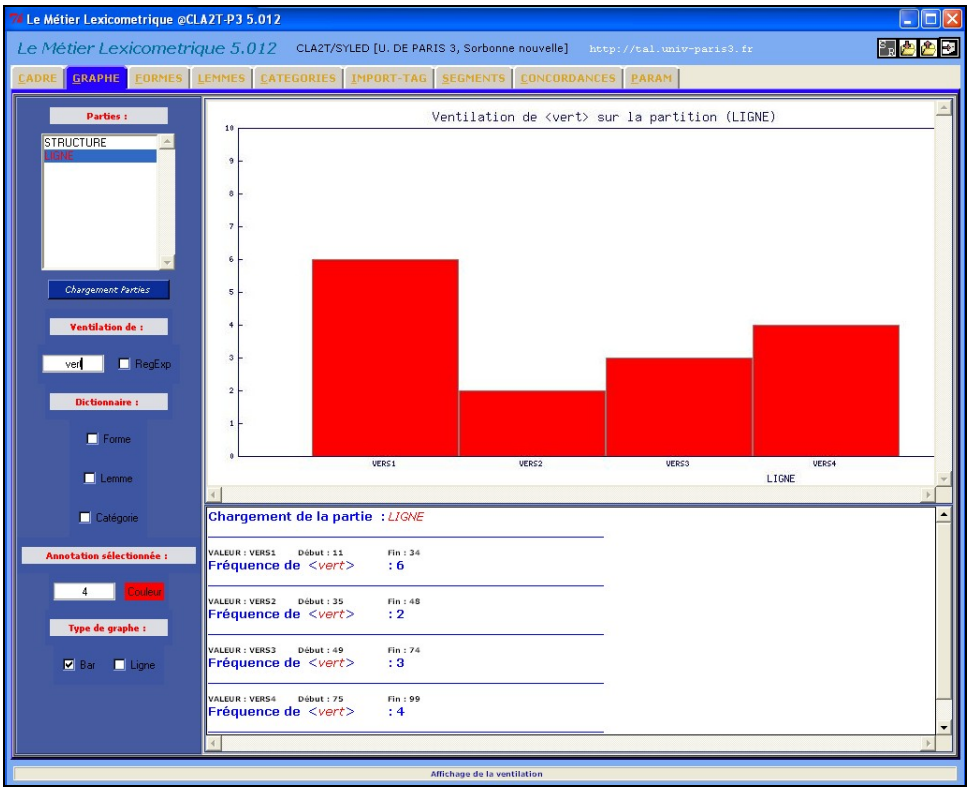


Figure 178 : *Le Trameur* ventilation d'item (annotation importée)

23 Ajout/modification d'annotation sur une *Trame* via les dictionnaires

La modification et l'insertion d'annotations a déjà été présentée via le *Gestionnaire de Sélection.* Ces 2 opérations sont aussi disponibles sur les contenus des dictionnaires.

Dans le ***Trameur***, chaque position de la ***Trame*** peut être associée à plusieurs systèmes d'annotation. On peut importer un système d'annotation sur l'intégralité des positions de la ***Trame*** (*cf* partie précédente). On peut aussi insérer ou modifier des annotations en sélectionnant des formes de la ***Trame*** dans le dictionnaire des formes. Dans l'exemple qui suit, on travaille sur le texte « Le Père Duchesne », on a paramétré le ***Trameur*** pour avoir un étiquetage du texte de travail. On a donc : le numéro d'annotation n°1 correspondant à « Forme », le numéro d'annotation n°2 correspondant à « Lemme », le numéro d'annotation n°3 correspondant à « Catégorie ». Toute annotation importée/insérée aura un numéro supérieur ou égal à 4.

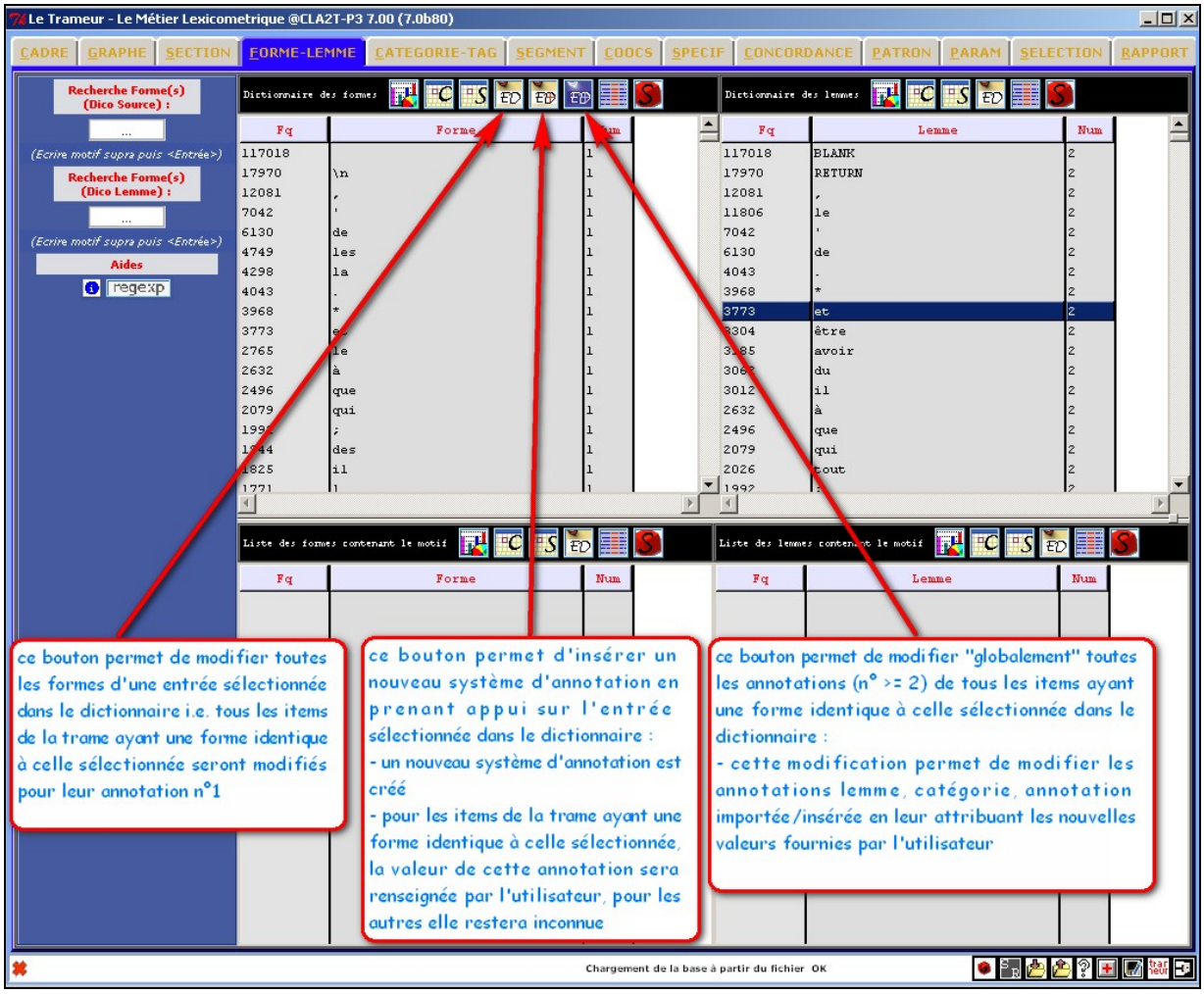
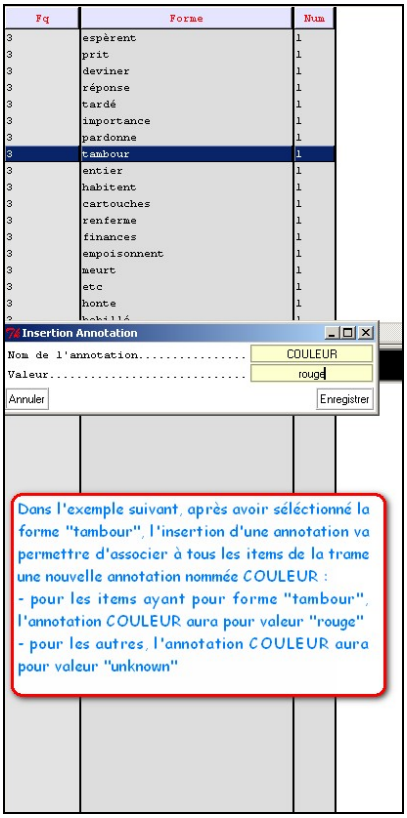
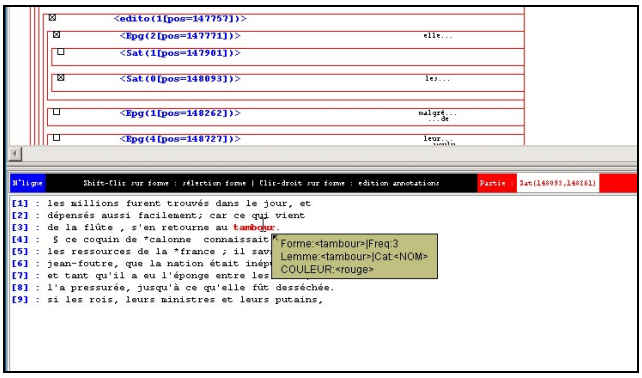

Le Trameur - Le Métier Lexicometrique @CLA2T-P3 7.00 (7.0b80)
CADRE
GRAPHE
SECTION
FORME-LEMME
CATEGORIE-TAG
SEGMENT
COOCS
SPECIF
CONCORDANCE
PATRON
PARAM
SELECTION
RAPPORT
Recherche Forme(s) (Dico Source) :
(Ecrire motif supra puis <Entrée>)
Recherche Forme(s) (Dico Lemme) :
(Ecrire motif supra puis <Entrée>)
Aides
regexp
Dictionnaire des formes
Fq
Forme
Num
Dictionnaire des lemmes
Fq
Lemme
Num
Liste des formes contenant le motif
Liste des lemmes contenant le motif
ce bouton permet de modifier toutes les formes d'une entrée sélectionnée dans le dictionnaire i.e. tous les items de la trame ayant une forme identique à celle sélectionnée seront modifiés pour leur annotation n°1
ce bouton permet d'insérer un nouveau système d'annotation en prenant appui sur l'entrée sélectionnée dans le dictionnaire :
- un nouveau système d'annotation est créé
- pour les items de la trame ayant une forme identique à celle sélectionnée, la valeur de cette annotation sera renseignée par l'utilisateur, pour les autres elle restera inconnue
ce bouton permet de modifier "globalement" toutes les annotations (n° >= 2) de tous les items ayant une forme identique à celle sélectionnée dans le dictionnaire :
- cette modification permet de modifier les annotations lemme, catégorie, annotation importée/insérée en leur attribuant les nouvelles valeurs fournies par l'utilisateur
Chargement de la base à partir du fichier OK

Figure 179 : *Le Trameur* édition de la *Trame* par les formes

23.1 Ajout d'annotation


Fq
Forme
Num
3 espèrent 1
3 prit 1
3 deviner 1
3 réponse 1
3 tardé 1
3 importance 1
3 pardonne 1
3 tambour 1
3 entier 1
3 habitent 1
3 cartouches 1
3 renferme 1
3 finances 1
3 empoisonnent 1
3 meurt 1
3 etc 1
3 honte 1
Insertion Annotation
Nom de l'annotation..... COULEUR
Valeur..... rouge
Annuler
Enregistrer
Dans l'exemple suivant, après avoir séléctionné la forme "tambour", l'insertion d'une annotation va permettre d'associer à tous les items de la trame une nouvelle annotation nommée COULEUR :
- pour les items ayant pour forme "tambour", l'annotation COULEUR aura pour valeur "rouge"
- pour les autres, l'annotation COULEUR aura pour valeur "unknown"

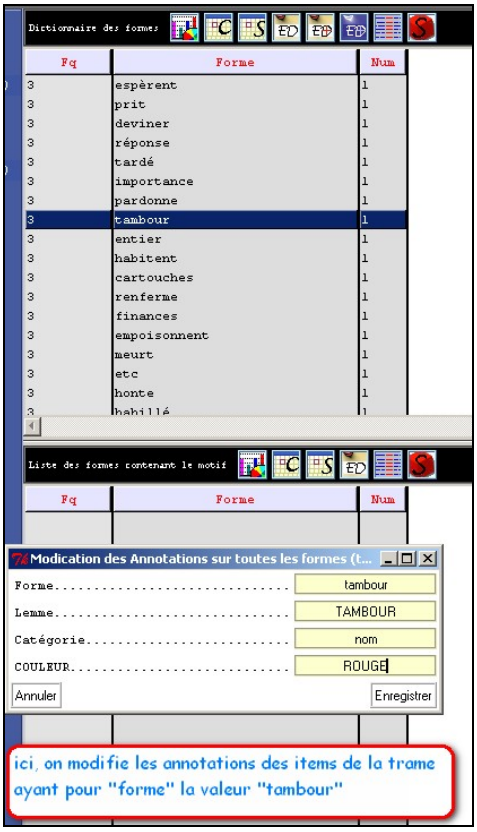

[1] : les millions furent trouvés dans le jour, et
[2] : dépensés aussi facilement; car ce qui vient
[3] : de la flûte , s'en retourne au tambour.
[4] : Ş ce coquin de "calonne connaissait
[5] : les ressources de la "france ; il sav
[6] : jean-foutre, que la nation était inép
[7] : et tant qu'il a eu l'éponge entre les
[8] : l'a pressurée, jusqu'à ce qu'elle fût desséchée.
[9] : si les rois, leurs ministres et leurs putains,
Forme:<tambour>|Freq:3
Lemme:<tambour>|Cat:<NOM>
COULEUR:<rouge>

Phase 1 : ajout d'une annotation

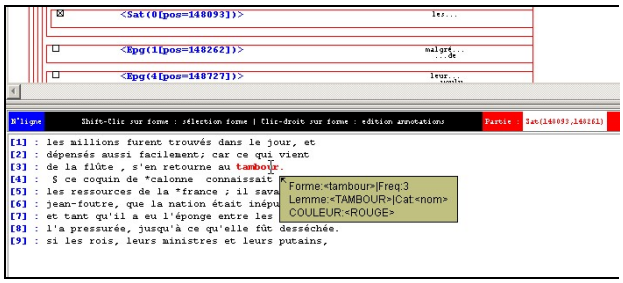
*Phase 2 : visualisation sur la **Trame***

Figure 180 : *Le Trameur* ajout d'annotation sur la *Trame* via une forme sélectionnée

23.2 Modification d'annotation



Phase 1 : ajout d'une annotation


[1] : les millions furent trouvés dans le jour, et
[2] : dépensés aussi facilement; car ce qui vient
[3] : de la flûte , s'en retourne au tambour.
[4] : $ ce coquin de *calonne connaissait
[5] : les ressources de la *france ; il sava
[6] : jean-foutre, que la nation était inépu
[7] : et tant qu'il a eu l'éponge entre les
[8] : l'a pressurée, jusqu'à ce qu'elle fût desséchée.
[9] : si les rois, leurs ministres et leurs putains,
Forme:<tambour>|Freq:3
Lemme:<TAMBOUR>|Cat:<nom>
COULEUR:<ROUGE>

*Phase 2 : visualisation sur la **Trame***

Figure 181 : *Le Trameur* modification d'annotation sur la *Trame* via une forme sélectionnée

Au cours de la modification des annotations d'un item sélectionné, si les valeurs des champs `Lemme`, `Catégorie` et `annotations importées` restent vides, les valeurs initiales seront respectivement conservées pour chacune des annotations concernées.

24 Des nuages de d'annotations sur la *Trame*

Ce module permet de générer un fichier (au format XML) donnant à voir une liste d'items de la ***Trame*** (*via* un niveau d'annotation choisi), chacun de ces items étant associé à l'ensemble des sections (zones de texte paramétrable) associées à cette annotation.

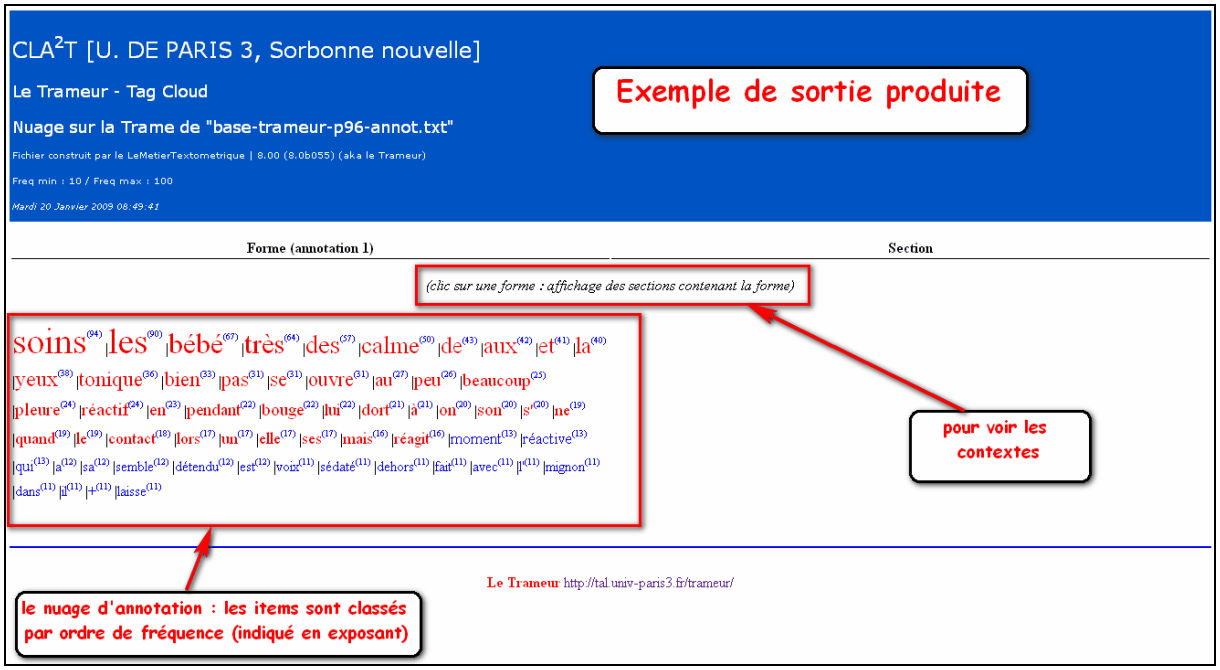

CLA²T [U. DE PARIS 3, Sorbonne nouvelle]
Le Trameur - Tag Cloud
Nuage sur la Trame de "base-trameur-p96-annot.txt"
Fichier construit par le LeMetierTextometrique | 8.00 (8.0b055) (aka le Trameur)
Freq min : 10 / Freq max : 100
Mardi 20 Janvier 2009 08:49:41
Exemple de sortie produite
Forme (annotation 1)
Section
(clic sur une forme : affichage des sections contenant la forme)
soins(94) |les(90) |bébé(67) |très(64) |des(57) |calme(50) |de(43) |aux(42) |et(41) |la(40)
|yeux(38) |tonique(36) |bien(33) |pas(31) |se(31) |ouvre(31) |au(27) |peu(26) |beaucoup(25)
|pleure(24) |réactif(24) |en(23) |pendant(22) |bouge(22) |lui(22) |dort(21) |à(21) |on(20) |son(20) |s'(20) |ne(19)
|quand(19) |le(19) |contact(18) |lors(17) |un(17) |elle(17) |ses(17) |mais(16) |réagit(16) |moment(13) |réactive(13)
|qui(13) |a(12) |sa(12) |semble(12) |détendu(12) |est(12) |voix(11) |sédaté(11) |dehors(11) |fait(11) |avec(11) |l'(11) |mignon(11)
|dans(11) |il(11) |+(11) |laisse(11)
pour voir les contextes
le nuage d'annotation : les items sont classés par ordre de fréquence (indiqué en exposant)
Le Trameur http://tal.univ-paris3.fr/trameur/

Figure 182 : *Le Trameur* un nuage d'annotations sur la *Trame*

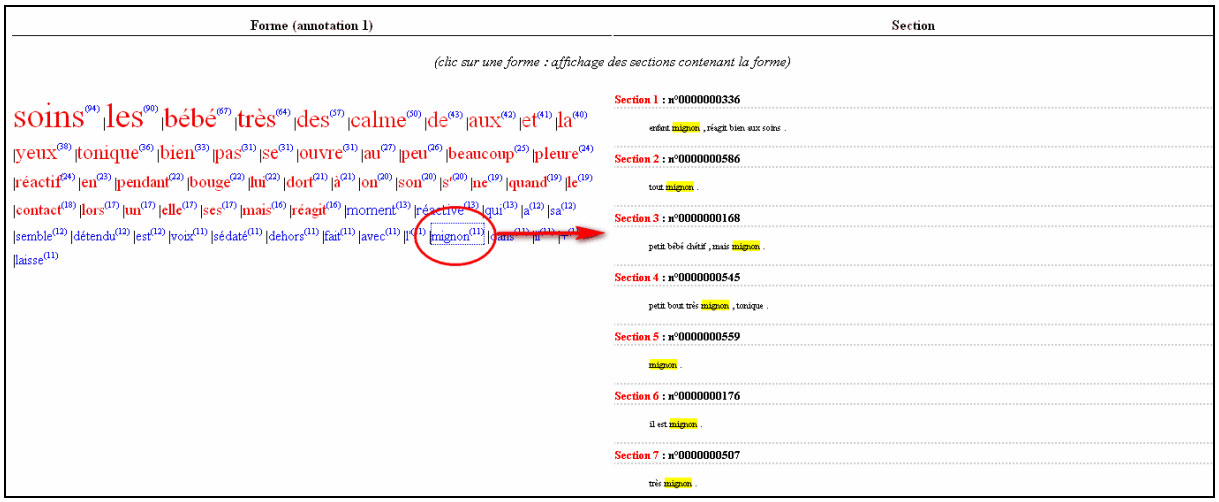

Forme (annotation 1)
Section
(clic sur une forme : affichage des sections contenant la forme)
soins(94) |les(90) |bébé(67) |très(64) |des(57) |calme(50) |de(43) |aux(42) |et(41) |la(40)
|yeux(38) |tonique(36) |bien(33) |pas(31) |se(31) |ouvre(31) |au(27) |peu(26) |beaucoup(25) |pleure(24)
|réactif(24) |en(23) |pendant(22) |bouge(22) |lui(22) |dort(21) |à(21) |on(20) |son(20) |s'(20) |ne(19) |quand(19) |le(19)
|contact(18) |lors(17) |un(17) |elle(17) |ses(17) |mais(16) |réagit(16) |moment(13) |réactive(13) |qui(13) |a(12) |sa(12)
|semble(12) |détendu(12) |est(12) |voix(11) |sédaté(11) |dehors(11) |fait(11) |avec(11) |l'(11) |mignon(11)
|laisse(11)
Section 1 : n°0000000336
enfant mignon , réagit bien aux soins .
Section 2 : n°0000000586
tout mignon .
Section 3 : n°0000000168
petit bébé chétif , mais mignon .
Section 4 : n°0000000545
petit bout très mignon , tonique .
Section 5 : n°0000000559
mignon .
Section 6 : n°0000000176
il est mignon .
Section 7 : n°0000000507
très mignon .

Figure 183 : *Le Trameur* un nuage d'annotations sur la *Trame* avec contextes

Le paramétrage du calcul de ce nuage se fait dans l'onglet **DIC**.

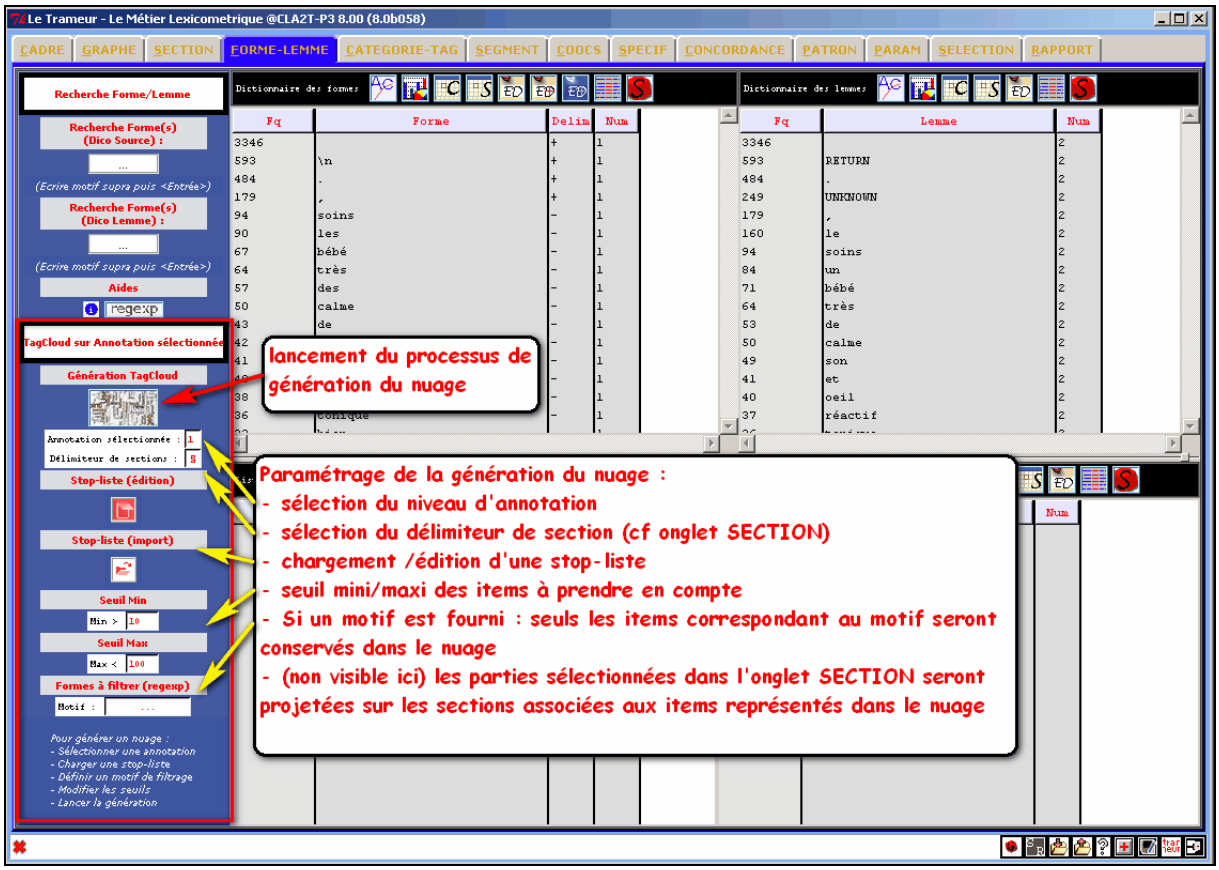


Figure 184 : *Le Trameur* paramétrage du nuage d'annotations

Des exemples de nuages d'annotations sur la ***Trame*** sont visibles sur la page web du ***Trameur***.

25 Relations de dépendance entre les items de *Trame* (via leurs annotations)

Cette partie introduit le traitement des relations de dépendances disponibles entre les items de la ***Trame***. Pour illustrer ces fonctionnalités, nous utilisons les bases textométriques suivantes (disponibles sur la page du logiciel) :

1. **Base "Corpus de contes français"** (Fairy Tales Corpus, FTC) construite à partir des ressources développées par Ismaïl El Maarouf. Chaque item de la ***Trame*** est associé à 7 niveaux d'annotation.

   SOURCES : I. El Maarouf et J. Villaneau (2012). *A French Fairy Tale Corpus syntactically and semantically annotated*. In Proceedings of LREC 2012. Istanbul (Tk).
   Pour les formes : copyright Momes.net, http://momes.net.
   Lemme et POS via treetagger : fichier paramètre langue (français, utf-8) développé par Christophe Benzitoun (ATILF)
   Pour les annotations : http://elmaarouf.legtux.org/Accueil.html

   Descriptif et sources des annotations :

   balise <f> (forme) : copyright Momes.net, http://elmaarouf.legtux.org/data/FTC_FORM.txt
   balise <l> (lemme) : fichier paramètre langue (français, utf-8) développé par Christophe Benzitoun (ATILF)
   balise <c> (POS) : fichier paramètre langue (français, utf-8) développé par Christophe Benzitoun (ATILF)
   Annotations complémentaires (http://elmaarouf.legtux.org/Accueil.html) :
   Identifiants de mots : http://elmaarouf.legtux.org/data/FTC_ID.txt (première balise <a>)
   • Annotation référentielle : http://elmaarouf.legtux.org/data/FTC_REF.txt (seconde balise <a>)
   • Annotation en rôles sémantiques pour 26 verbes : http://elmaarouf.legtux.org/data/FTC_SEMROLE.txt (troisième balise <a>)
   • Annotation syntaxique en dépendance pour 137 verbes : http://elmaarouf.legtux.org/data/FTC_SYNROLE.txt (dernière balise <a>)
   L'ensemble des annotations du corpus initial (+ lemme et POS) regroupées ici au format d'une base textométrique pour le ***Trameur***.

2. **Base "Rhapsodie2Trameur"** construite à partir des ressources développées dans le cadre du projet **Rhapsodie**. Chaque item de la *Trame* est associé à 13 niveaux d'annotation.

   SOURCES : projet Rhapsodie

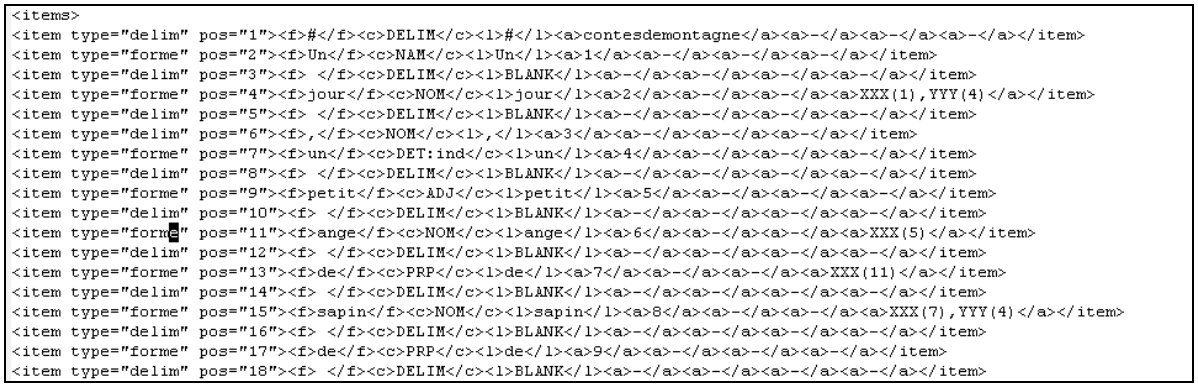
   Fichiers de codage microsyntaxique version béta 10/13 (zip)

   Descriptif et sources des annotations :
   Présentation du processus de transcodage des annotations Rhapsodie pour construire une base Textométrique (pdf)

3. **Base "Rhapsodie2Trameur" (v4)** construite à partir des ressources développées dans le cadre du projet **Rhapsodie**. Chaque item de la *Trame* est associé à 38 niveaux d'annotation (prosodie, micro et macro-syntaxe).

25.1 Extrait de la Trame : « un jour, un petit ange de sapin de … »

Dans les exemples présentés *infra*, nous allons travailler sur un extrait de la base des contes dont l'image suivante donne à voir le début de la ***Trame*** telle qu'elle est représentée dans la base textométrique disponible :



```
<items>
<item type="delim" pos="1"><f>#</f><c>DELIM</c><l>#</l><a>contesdemontagne</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="2"><f>Un</f><c>NAM</c><l>Un</l><a>1</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="3"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="4"><f>jour</f><c>NOM</c><l>jour</l><a>2</a><a>-</a><a>-</a><a>XXX(1),YYY(4)</a></item>
<item type="delim" pos="5"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="6"><f>,</f><c>NOM</c><l>,</l><a>3</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="7"><f>un</f><c>DET:ind</c><l>un</l><a>4</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="8"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="9"><f>petit</f><c>ADJ</c><l>petit</l><a>5</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="10"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="11"><f>ange</f><c>NOM</c><l>ange</l><a>6</a><a>-</a><a>-</a><a>XXX(5)</a></item>
<item type="delim" pos="12"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="13"><f>de</f><c>PRP</c><l>de</l><a>7</a><a>-</a><a>-</a><a>XXX(11)</a></item>
<item type="delim" pos="14"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="15"><f>sapin</f><c>NOM</c><l>sapin</l><a>8</a><a>-</a><a>-</a><a>XXX(7),YYY(4)</a></item>
<item type="delim" pos="16"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="17"><f>de</f><c>PRP</c><l>de</l><a>9</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="18"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a></item>
```

Cet extrait met en avant des relations de dépendances sur plusieurs items de la ***Trame*** :

- L'item en position 4 (dont la forme est « jour ») porte dans son annotation n°7 la valeur :

`XXX(1), YYY(4)`

Cette information traduit la présence de 2 relations de dépendance pour cet item :

- une nommée XXX, l'autre YYY ; la première associe l'item courant avec un item dont l'identifiant est 1,
- la seconde associe l'item courant avec un item dont l'identifiant est 4.

Dans cette base, l'identification des éléments de la ***Trame*** ne coïncide pas avec leur position sur la ***Trame***, une annotation particulière (la première balise <a> correspondant à l'annotation n°4) porte cette information.

On a donc :

« jour » (position 4) → « Un » (position 2, identifiant 1) par la relation XXX
« jour » (position 4) → « un » (position 7, identifiant 4) par la relation YYY

25.2 Recherche de relations

L'onglet **GRAPHE** permet de lancer une requête de recherche de relations entre items de la ***Trame***.

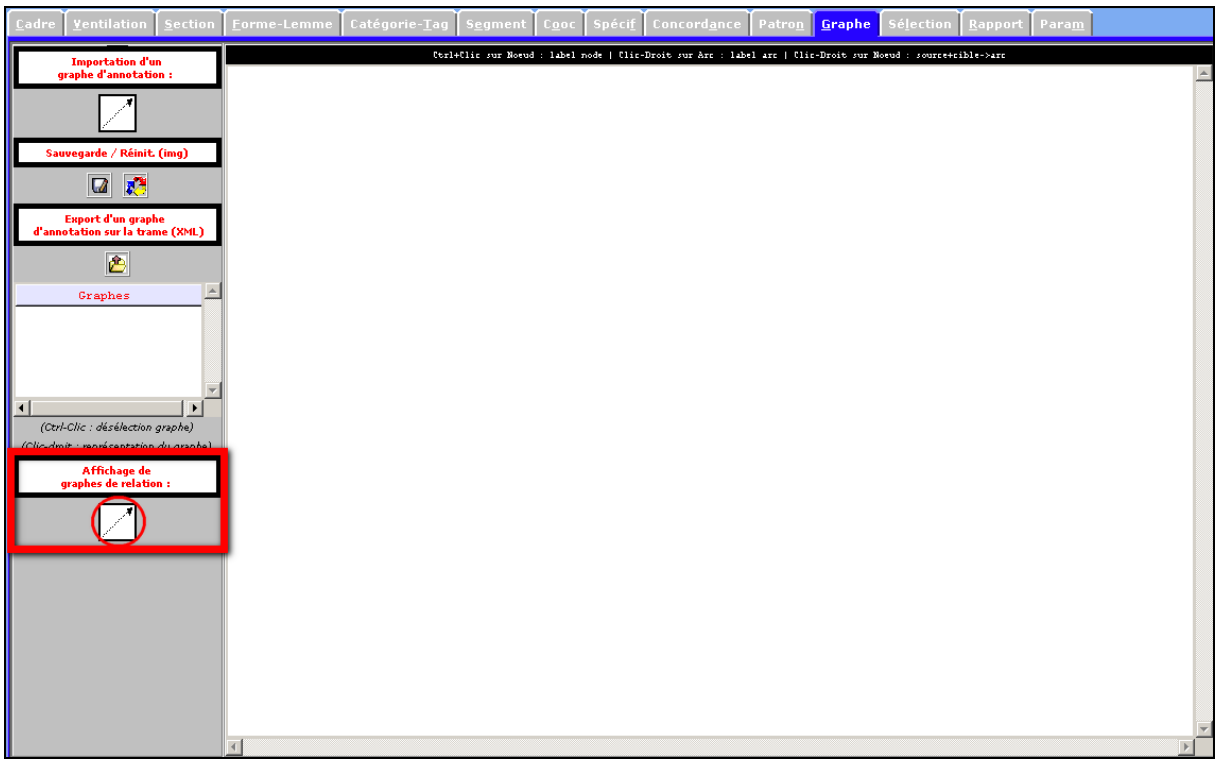


Figure 185 : ***Le Trameur*** **recherche de relations de dépendance via l'onglet GRAPHE**

25.2.1 Sélection des relations à afficher

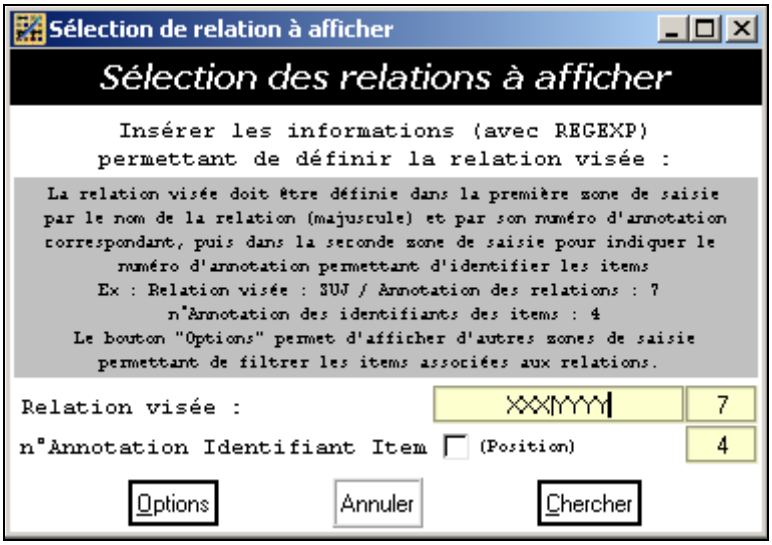

Sélection de relation à afficher
Sélection des relations à afficher
Insérer les informations (avec REGEXP)
permettant de définir la relation visée :
La relation visée doit être définie dans la première zone de saisie
par le nom de la relation (majuscule) et par son numéro d'annotation
correspondant, puis dans la seconde zone de saisie pour indiquer le
numéro d'annotation permettant d'identifier les items
Ex : Relation visée : SUJ / Annotation des relations : 7
n°Annotation des identifiants des items : 4
Le bouton "Options" permet d'afficher d'autres zones de saisie
permettant de filtrer les items associées aux relations.
Relation visée :
XXX|YYY
7
n°Annotation Identifiant Item
(Position)
4
Options
Annuler
Chercher

Figure 186 : ***Le Trameur*** **paramétrage de la recherche de dépendance sur la** ***Trame***

Pour afficher les relations disponibles dans la base, un menu permet de définir ces relations.

Dans l'exemple ci-dessus, on définit le modèle « XXX|YYY » permettant de représenter soit la relation XXX soit la relation YYY en spécifiant le numéro de l'annotation portant les relations de dépendance à scruter.

On précise aussi ici où repérer les identifications des items concernés (annotation n°4).

Dans la base traitée ici, chaque item de la ***Trame*** est associé à un niveau d'annotation (numéro 4) portant une valeur permettant d'identifier ces items. Si ce niveau d'annotation n'est pas disponible, il est possible d'utiliser une « niveau d'annotation virtuel » pour identifier les items : celui fourni par la position des items sur la ***Trame***. Pour activer cette identification des items par leur position, il suffit de cocher la case disponible « (Position) ».

25.2.2 Graphe des relations

L'activation de la recherche produit le graphe des relations disponibles sur le modèle visé :

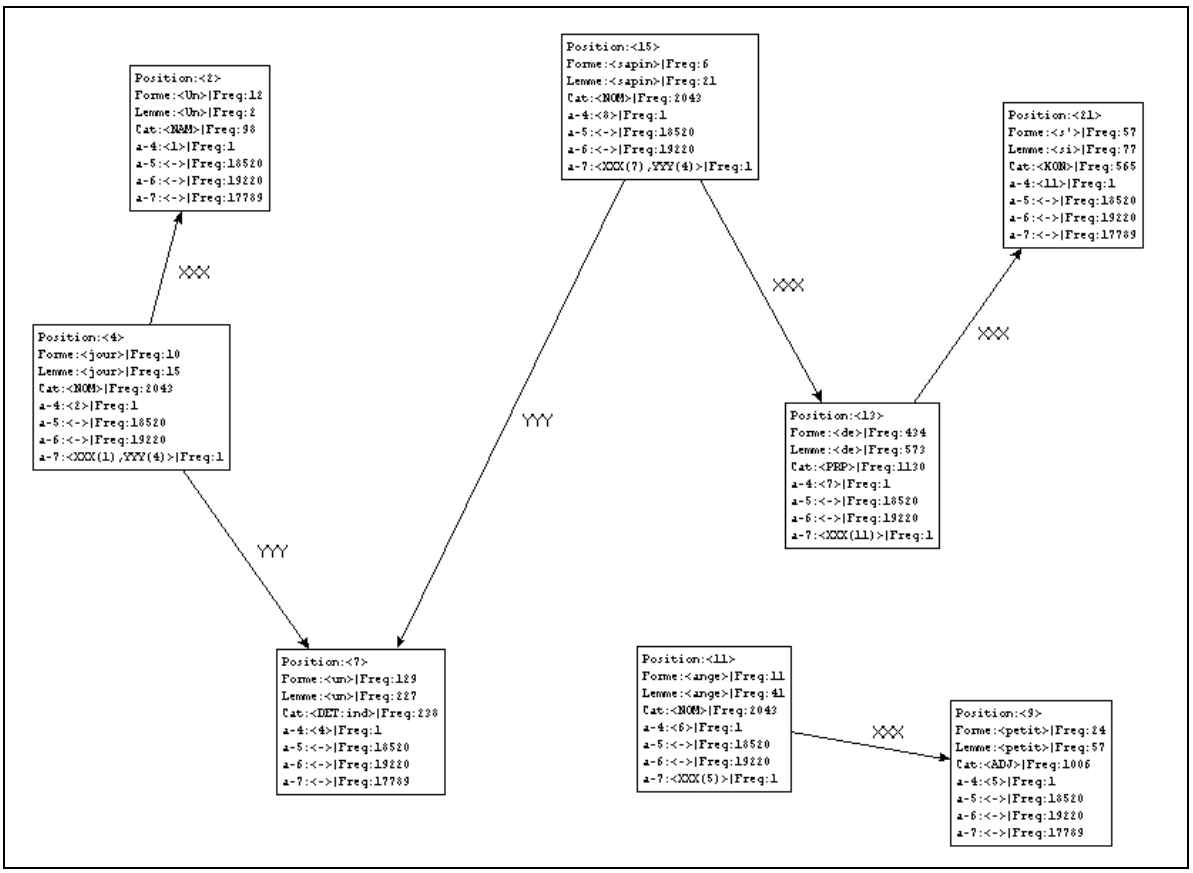


Figure 187 : *Le Trameur* affichage du graphe des dépendances

25.3 Sélection des relations avec filtrage sur le nœud source de la relation

On peut affiner le modèle de recherche de relations en filtrant sur les annotations associées aux nœuds des relations visées. Dans l'exemple suivant, on s'intéresse aux objets (relation OBJ) des items contenant la chaîne de caractère « connu » au niveau de l'annotation n°1 (forme) :

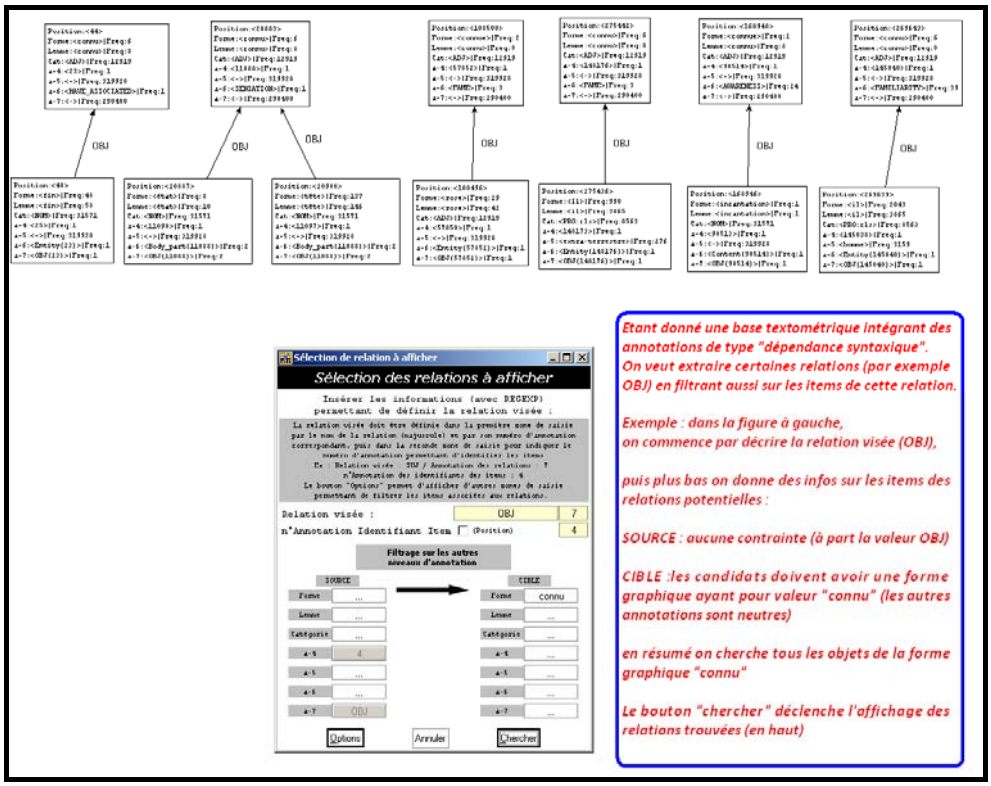

Sélection de relation à afficher
Sélection des relations à afficher
Relation visée :
OBJ
Filtrage sur les autres niveaux d'annotation
SOURCE
CIBLE
Forme
Lemme
Catégorie
connu
Options
Annuler
Chercher
Etant donné une base textométrique intégrant des annotations de type "dépendance syntaxique". On veut extraire certaines relations (par exemple OBJ) en filtrant aussi sur les items de cette relation.
Exemple : dans la figure à gauche, on commence par décrire la relation visée (OBJ),
puis plus bas on donne des infos sur les items des relations potentielles :
SOURCE : aucune contrainte (à part la valeur OBJ)
CIBLE :les candidats doivent avoir une forme graphique ayant pour valeur "connu" (les autres annotations sont neutres)
en résumé on cherche tous les objets de la forme graphique "connu"
Le bouton "chercher" déclenche l'affichage des relations trouvées (en haut)

Figure 188 : *Le Trameur* paramétrage de la recherche de dépendance avec filtrage

Autre illustration sur la base « Corpus de conte » sur lequel on a cherché la relation « SUJ|OBJ », le graphique donne à voir un extrait du graphe produit :

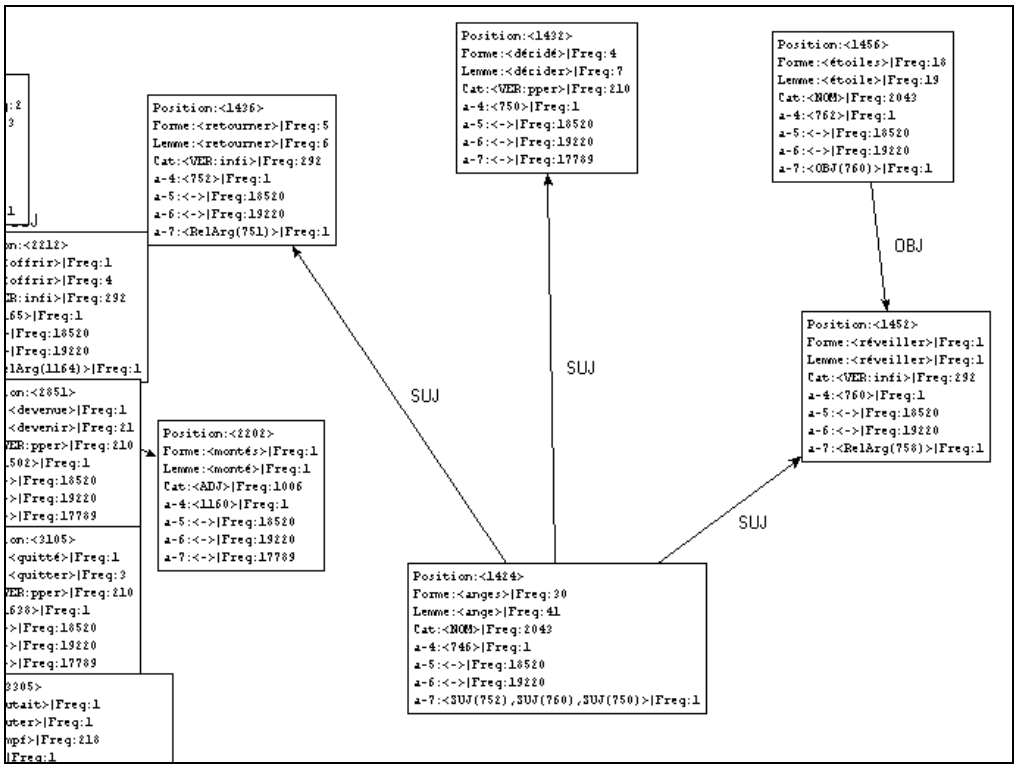


Figure 189 : *Le Trameur* affichage du graphe des dépendances (1)

De même, sur la base *Rhapsodie2Trameur*, la figure suivante illustre la recherche des « objets » du lemme « affirmer » :

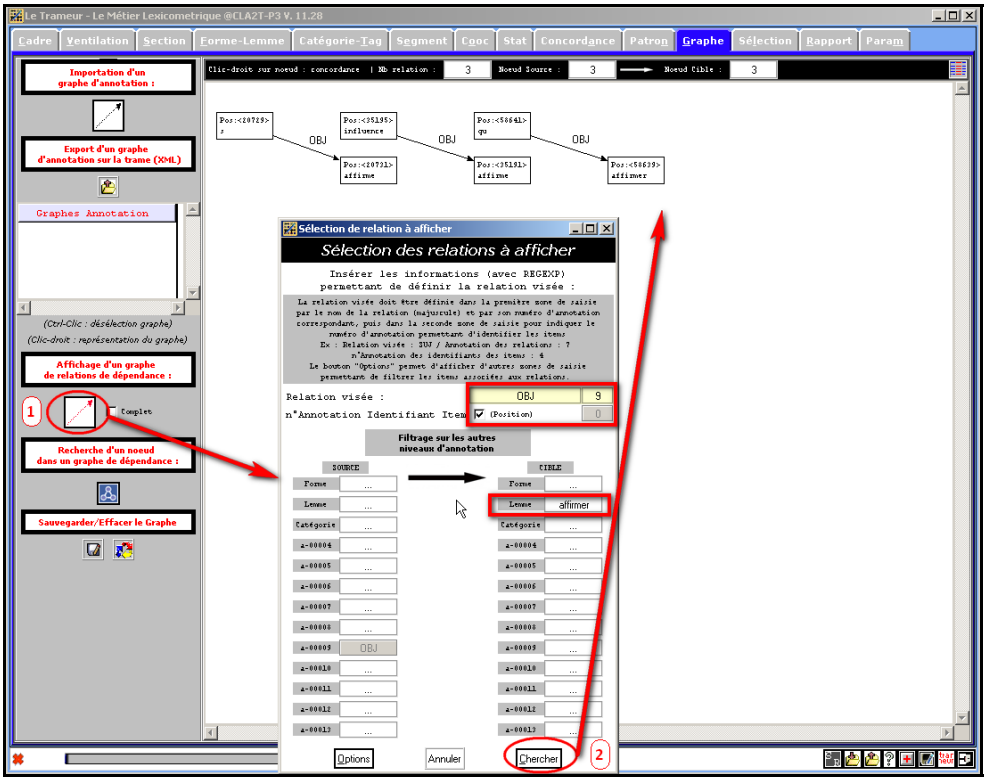


Figure 190 : *Le Trameur* affichage du graphe des dépendances (2)

Dans la figure précédente, on cherche des relations de dépendance de type OBJ en contraignant un des composants de la relation : il doit porter le lemme "affirmer" (i.e. on cherche les objets d'affirmer). Dans la figure suivante, la relation cherchée est double (SUB ou OBJ), la cible de la relation impose une valeur pour le lemme (« penser ») et pour l'annotation n°9 (ROOT)

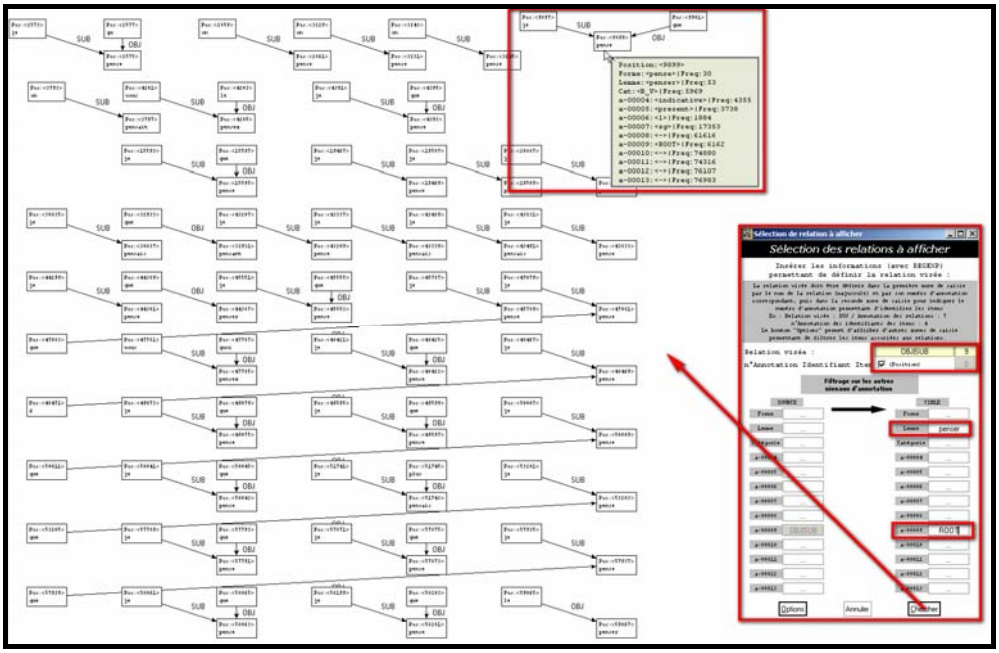
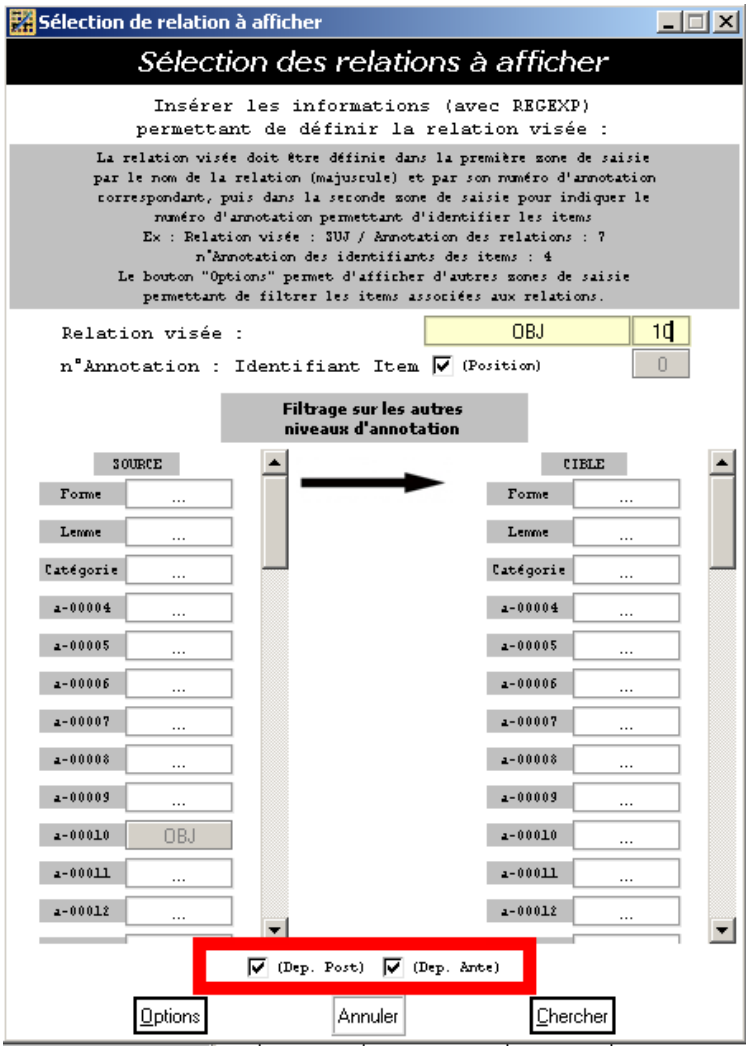


Figure 191 : *Le Trameur* affichage du graphe des dépendances (3)

Dans les 2 graphes précédents, l'affichage des nœuds est celui produit par défaut : affichage de la position de l'item sur la ***Trame*** et de sa forme graphique. On peut visualiser l'ensemble des annotations d'un nœud donné via le mode Aperçu disponible dans tous les éditeurs (en passant la souris au-dessus de l'objet visé) : ci-dessus, l'item en position 9899 (lemme : penser) est mis au jour. Dans les précédentes figures, l'affichage d'un nœud est directement rendu « complet » via l'option du même nom disponible et activable avant de charger un graphe.

25.4 Sélection des relations avec filtrage sur la position du dépendant

Une option supplémentaire permet de filtrer la recherche de relation en contraignant la position du dépendant : postposition ou antéposition.



Par défaut, la recherche met au jour les dépendants antéposés et postposés, si on décoche une des options, la recherche ne s'appliquera que sur les dépendants correspondant à l'option cochée.

25.5 Actions sur les nœuds du graphe : retour en contexte

Dans l'exemple qui suit, on s'intéresse aux relations `SUJ` ou `OBJ` concernant les items ayant pour forme « anges » dans la base des contes :

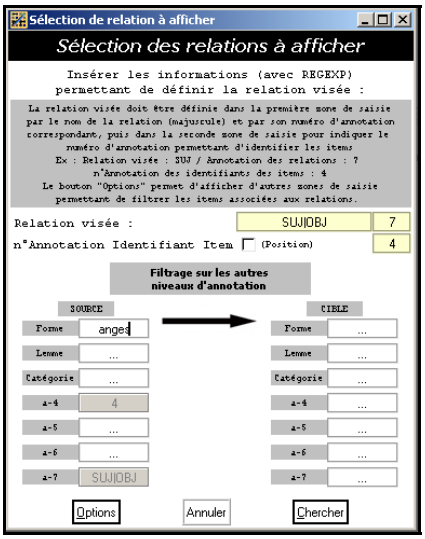

Sélection de relation à afficher
Sélection des relations à afficher
Insérer les informations (avec REGEXP)
permettant de définir la relation visée :
La relation visée doit être définie dans la première zone de saisie par le nom de la relation (majuscule) et par son numéro d'annotation correspondant, puis dans la seconde zone de saisie pour indiquer le numéro d'annotation permettant d'identifier les items
Ex : Relation visée : SUJ / Annotation des relations : 7
n°Annotation des identifiants des items : 4
Le bouton "Options" permet d'afficher d'autres zones de saisie permettant de filtrer les items associés aux relations.
Relation visée :
SUJ|OBJ
7
n°Annotation Identifiant Item
(Position)
4
Filtrage sur les autres niveaux d'annotation
SOURCE
CIBLE
Forme
anges
Lemme
Catégorie
a-4
4
a-5
a-6
a-7
SUJ|OBJ
Options
Annuler
Chercher

Figure 192 : *Le Trameur* paramétrage de la recherche de dépendance avec filtrage

Le graphe produit a l'allure suivante :

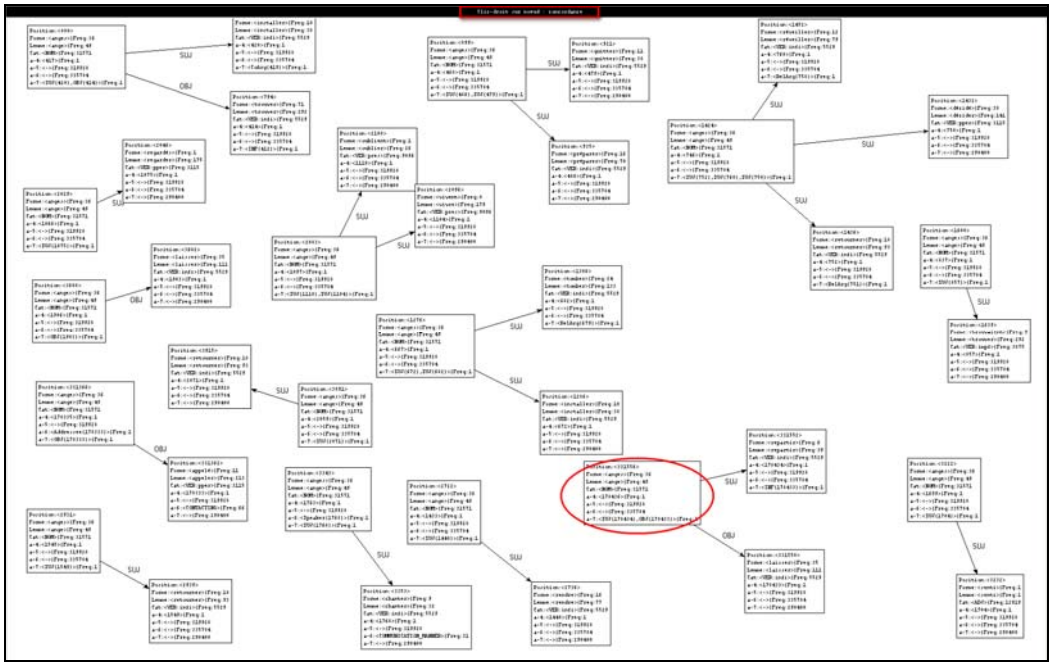


Figure 193 : *Le Trameur* affichage du graphe des dépendances

La combinaison de touche « clic-droit » sur un nœud du graphe permet d'afficher une concordance de l'item visé :

```
------------------------------PARTIE{TITRE=L'oiseau parfait}-----------------------------
une nouvelle vie à un animal .Mais Dieu ne voulait pas laisser repartir ses anges comme ça !Alors ,soudain ,il eut une idée .-Qu' aimeriez -vous ,
```

Figure 194 : *Le Trameur* affichage de la concordance sur un des nœuds du graphe des dépendances

L'ensemble des nœuds peut aussi être projeté globalement sur le concordancier. Dans la figure suivante, on reprend le graphe produit sur la base *Rhapsodie2Trameur* montrant le résultat de la recherche des « objets » du lemme « affirmer » :

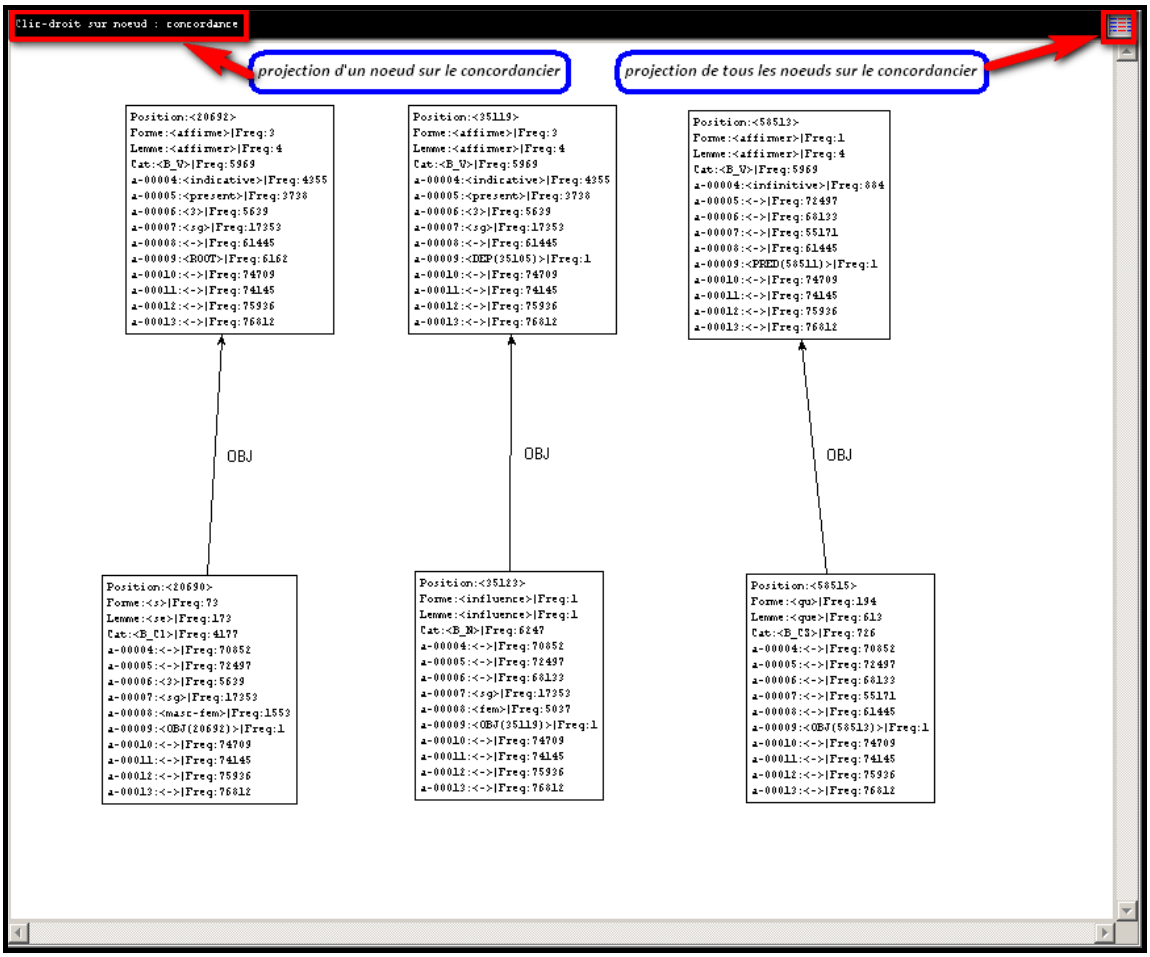


Figure 195 : *Le Trameur* affichage de la concordance sur tous les nœuds du graphe des dépendances

La concordance produite après projection de tous les noeuds a l'allure suivante :

```
------------------------------PARTIE{PARTIE=M2004}------------------------------
                $ un nouvel ordre international s'affirme peu à peu
------------------------------PARTIE{PARTIE=M2001}------------------------------
                qu'une nation affirme son influence qu'elle pèse dans
------------------------------PARTIE{PARTIE=D2013}------------------------------
                  Rachida Dati pour oser affirmer qu'il ne s'agit
```

Figure 196 : *Le Trameur* Concordance sur tous les nœuds du graphe des dépendances (1)

Les sources de la relation sont coloriées en rouge (pôle de la concordance), les cibles en vert (i.e ici le verbe est en vert, son objet en rouge). On peut aussi varier l'affichage des annotations dans la concordance en matérialisant par exemple la catégorie de chaque item :

```
------------------------------PARTIE{PARTIE=M2004}------------------------------
  $ un-B_D  nouvel-B_Adj  ordre-B_N  international-B_Adj  s-B_Cl '-I_Cl affirme-B_V peu-B_Adv  à-I_Adv  peu-I_Adv
------------------------------PARTIE{PARTIE=M2001}------------------------------
qu-B_Qu '-I_Qu une-B_D  nation-B_N  affirme-B_V son-B_D  influence-B_N  qu-B_Qu '-I_Qu elle-B_Cl  pèse-B_V  dans-B_Pre
------------------------------PARTIE{PARTIE=D2013}------------------------------
  Rachida-B_N  Dati-B_N  pour-B_Pre  oser-B_V  affirmer-B_V  qu-B_CS '-I_CS il-B_Cl  ne-B_Cl  s-B_Cl '-I_Cl agit-B_V
```

Figure 197 : *Le Trameur* Concordance sur tous les nœuds du graphe des dépendances (2)

Toutes les zones d'édition du ***Trameur*** permettent de mettre au jour les annotations de la ***Trame*** (cf « Marquage des annotations de la *Trame* »). Si on considère la figure suivante, elle présente un extrait de la concordance construite à partir du graphe ayant permis d'extraire la relation `OBJ` :

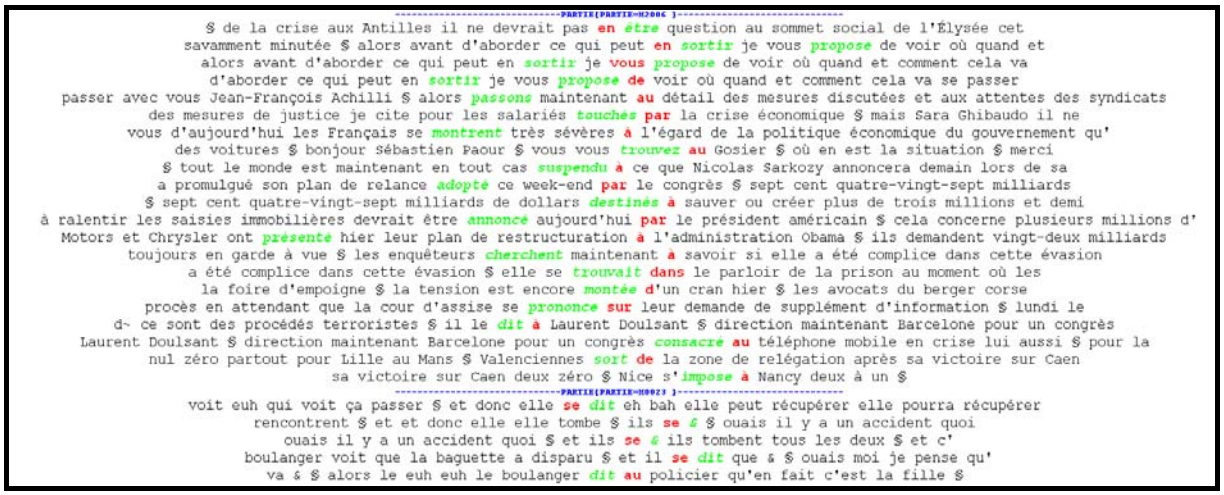

$ de la crise aux Antilles il ne devrait pas en être question au sommet social de l'Élysée cet
savamment minutée $ alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et
alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va
d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va se passer
passer avec vous Jean-François Achilli $ alors passons maintenant au détail des mesures discutées et aux attentes des syndicats
des mesures de justice je cite pour les salariés touchés par la crise économique $ mais Sara Ghibaudo il ne
vous d'aujourd'hui les Français se montrent très sévères à l'égard de la politique économique du gouvernement qu'
des voitures $ bonjour Sébastien Paour $ vous vous trouvez au Gosier $ où en est la situation $ merci
$ tout le monde est maintenant en tout cas suspendu à ce que Nicolas Sarkozy annoncera demain lors de sa
a promulgué son plan de relance adopté ce week-end par le congrès $ sept cent quatre-vingt-sept milliards
$ sept cent quatre-vingt-sept milliards de dollars destinés à sauver ou créer plus de trois millions et demi
à ralentir les saisies immobilières devrait être annoncé aujourd'hui par le président américain $ cela concerne plusieurs millions d'
Motors et Chrysler ont présenté hier leur plan de restructuration à l'administration Obama $ ils demandent vingt-deux milliards
toujours en garde à vue $ les enquêteurs cherchent maintenant à savoir si elle a été complice dans cette évasion
a été complice dans cette évasion $ elle se trouvait dans le parloir de la prison au moment où les
la foire d'empoigne $ la tension est encore montée d'un cran hier $ les avocats du berger corse
procès en attendant que la cour d'assise se prononce sur leur demande de supplément d'information $ lundi le
d~ ce sont des procédés terroristes $ il le dit à Laurent Doulsant $ direction maintenant Barcelone pour un congrès
Laurent Doulsant $ direction maintenant Barcelone pour un congrès consacré au téléphone mobile en crise lui aussi $ pour la
nul zéro partout pour Lille au Mans $ Valenciennes sort de la zone de relégation après sa victoire sur Caen
sa victoire sur Caen deux zéro $ Nice s'impose à Nancy deux à un $
voit euh qui voit ça passer $ et donc elle se dit eh bah elle peut récupérer elle pourra récupérer
rencontrent $ et et donc elle elle tombe $ ils se & $ ouais il y a un accident quoi
ouais il y a un accident quoi $ et ils se & ils tombent tous les deux $ et c'
boulanger voit que la baguette a disparu $ et il se dit que & $ ouais moi je pense qu'
va & $ alors le euh euh le boulanger dit au policier qu'en fait c'est la fille $

Figure 198 : *Le Trameur* Coloration d'annotation dans le concordancier (1)

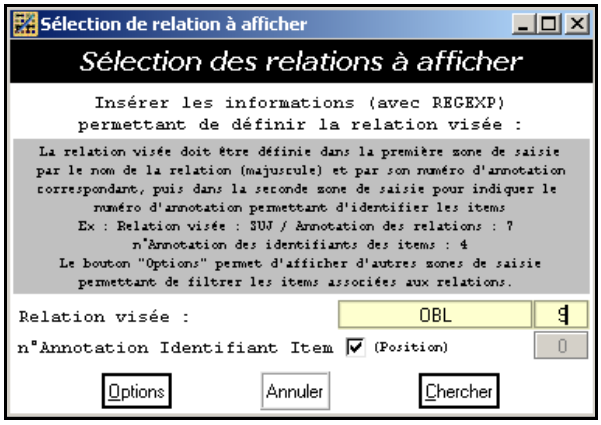
Le marquage de certains items est possible en sélectionnant les annotations à mettre au jour :

Sélection des annotations à colorier...
Annuler
Chercher
$ de la crise aux Antilles il ne devrait pas en être question au sommet social de l'Élysée cet
savamment minutée $ alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et
alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va
d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va se passer
passer avec vous Jean-François Achilli $ alors passons maintenant au détail des mesures discutées et aux attentes des syndicats
des mesures de justice je cite pour les salariés touchés par la crise économique $ mais Sara Ghibaudo il ne
vous d'aujourd'hui les Français se montrent très sévères à l'égard de la politique économique du gouvernement qu'
des voitures $ bonjour Sébastien Paour $ vous vous trouvez au Gosier $ où en est la situation $ merci
$ tout le monde est maintenant en tout cas suspendu à ce que Nicolas Sarkozy annoncera demain lors de sa
a promulgué son plan de relance adopté ce week-end par le congrès $ sept cent quatre-vingt-sept milliards
$ sept cent quatre-vingt-sept milliards de dollars destinés à sauver ou créer plus de trois millions et demi
à ralentir les saisies immobilières devrait être annoncé aujourd'hui par le président américain $ cela concerne plusieurs millions
Motors et Chrysler ont présenté hier leur plan de restructuration à l'administration Obama $ ils demandent vingt-deux milliards
toujours en garde à vue $ les enquêteurs cherchent maintenant à savoir si elle a été complice dans cette évasion
a été complice dans cette évasion $ elle se trouvait dans le parloir de la prison au moment où les
la foire d'empoigne $ la tension est encore montée d'un cran hier $ les avocats du berger corse
procès en attendant que la cour d'assise se prononce sur leur demande de supplément d'information $ lundi le
d~ ce sont des procédés terroristes $ il le dit à Laurent Doulsant $ direction maintenant Barcelone pour un congrès
Laurent Doulsant $ direction maintenant Barcelone pour un congrès consacré au téléphone mobile en crise lui aussi $ pour la
nul zéro partout pour Lille au Mans $ Valenciennes sort de la zone de relégation après sa victoire sur Caen
sa victoire sur Caen deux zéro $ Nice s'impose à Nancy deux à un $
it euh qui voit ça passer $ et donc elle se dit eh bah elle peut récupérer elle pourra récupérer
rencontrent $ et et donc elle elle tombe $ ils se & $ ouais il y a un accident quoi
ouais il y a un accident quoi $ et ils se & ils tombent tous les deux $ et c'
boulanger voit que la baguette a disparu $ et il se dit que & $ ouais moi je pense qu'
va & $ alors le euh euh le boulanger dit au policier qu'en fait c'est la fille $
j'accorde une puissance énorme à l'acte d'écrire $ mais comme toujours l'
des moments où on écrit parce qu'on pense participer à un combat $ ça a été le cas disons
rivain ou d'écrivant $ et puis peu à peu se dégage finalement euh la vérité une vérité plus plus
est très très importante $ ça serait assez compliqué d'y répondre $ mais d'abord nous savons bien
très très importante $ ça serait assez compliqué d'y répondre $ mais d'abord nous savons bien qu'
à pas de pensée préalable au langage qui est en nous finalement euh $ et par conséquent euh &
cas comme le mien euh penser euh est immédiatement lié à une forme que je si je puis dire je

Figure 199 : *Le Trameur* Coloration d'annotation dans le concordancier (2)

25.6 Rechercher dans un graphe de dépendance

On présente tout d'abord le résultat produit par la requête suivante « recherche de la relation OBL » :


Sélection de relation à afficher
Sélection des relations à afficher
Insérer les informations (avec REGEXP)
permettant de définir la relation visée :
La relation visée doit être définie dans la première zone de saisie par le nom de la relation (majuscule) et par son numéro d'annotation correspondant, puis dans la seconde zone de saisie pour indiquer le numéro d'annotation permettant d'identifier les items
Ex : Relation visée : SUJ / Annotation des relations : 7
n°Annotation des identifiants des items : 4
Le bouton "Options" permet d'afficher d'autres zones de saisie permettant de filtrer les items associées aux relations.
Relation visée :
OBL
9
n°Annotation Identifiant Item
(Position)
0
Options
Annuler
Chercher

Le résultat produit a l'allure suivante :

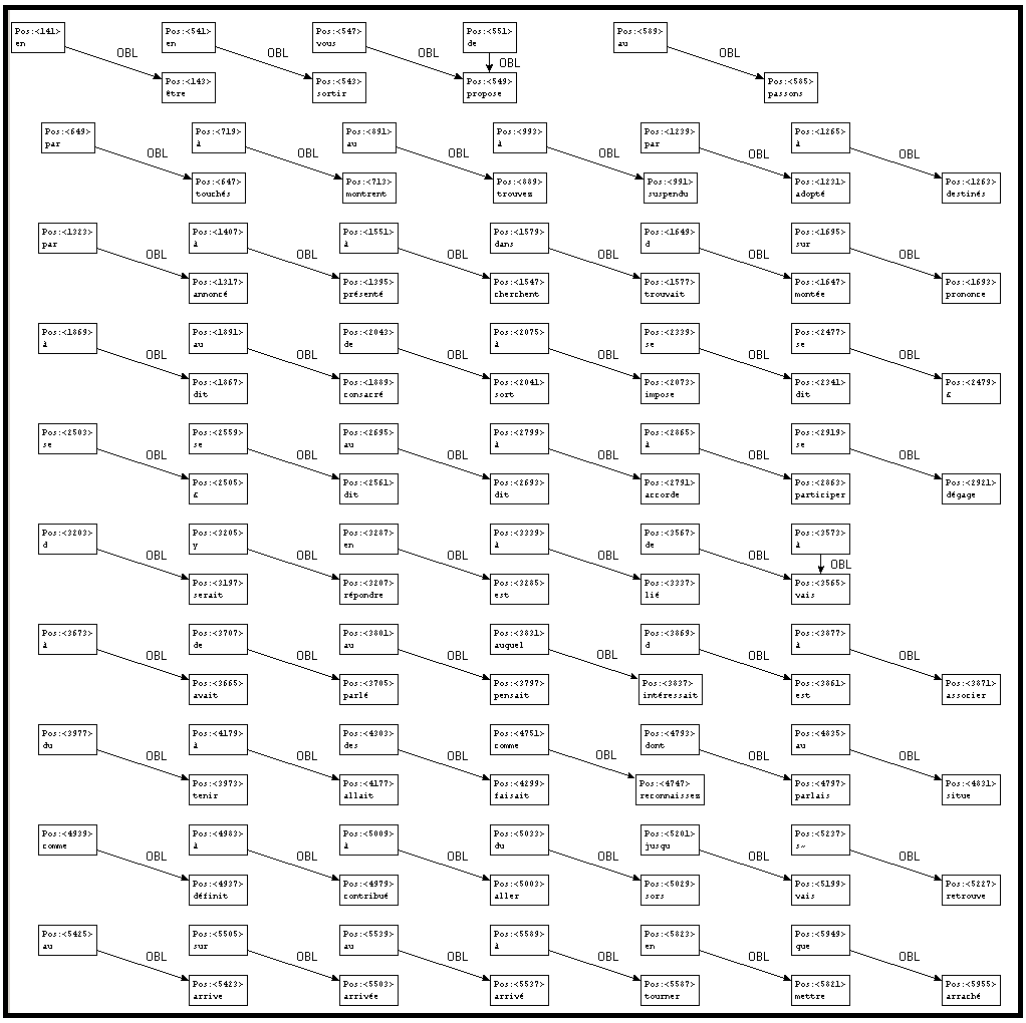


Figure 200 : *Le Trameur* Graphe de la relation OBL

On peut mettre au jour dans ce graphe certains nœuds en filtrant leurs annotations. Le processus de filtrage des nœuds du graphe permet de sélectionner certains nœuds sur la base

des différentes annotations disponibles. Dans l'exemple suivant, on veut mettre au jour les nœuds de catégorie B_V :

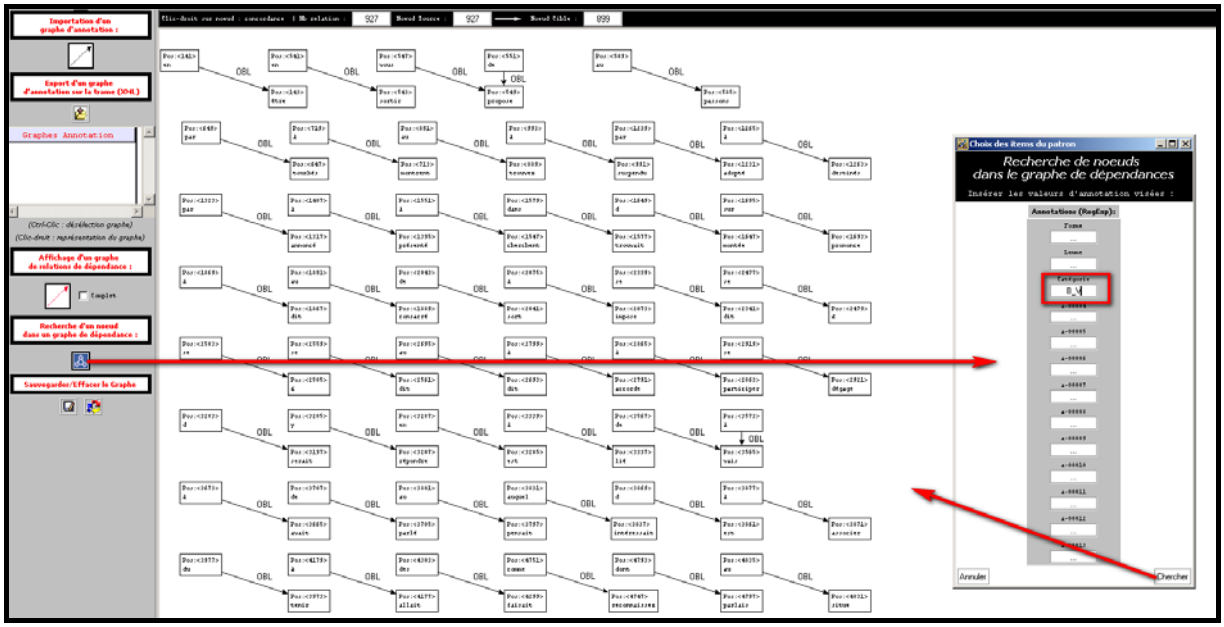


Figure 201 : *Le Trameur* Recherche dans un graphe de relation (1)

Le résultat produit a l'allure suivante :

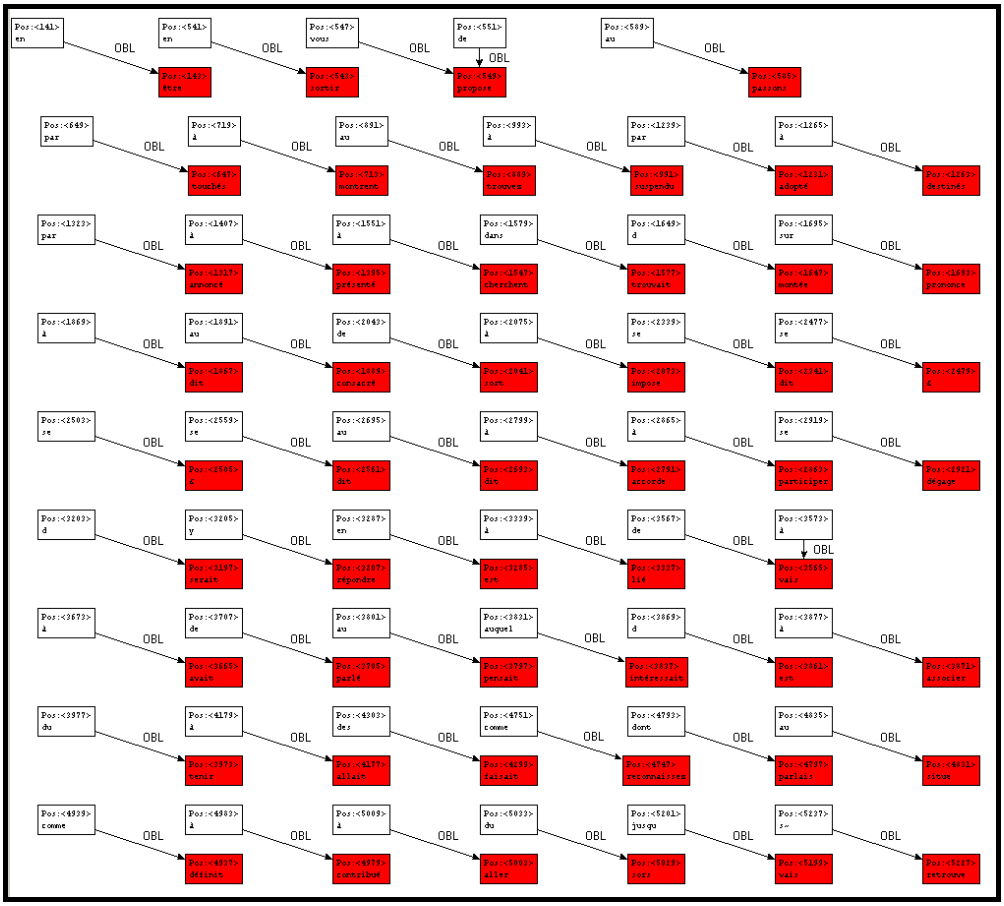


Figure 202 : *Le Trameur* Recherche dans un graphe de relation (2)

Dans la figure qui suit, 3 filtrages successifs sont réalisés pour marquer successivement les nœuds ayant pour catégorie `B_V`, puis `B_N` et enfin `B_Pre` ; chaque requête est précédée par la modification de la couleur à utiliser pour le marquage des nœuds visés (verbe en bleu, nom en vert et préposition en orange) :

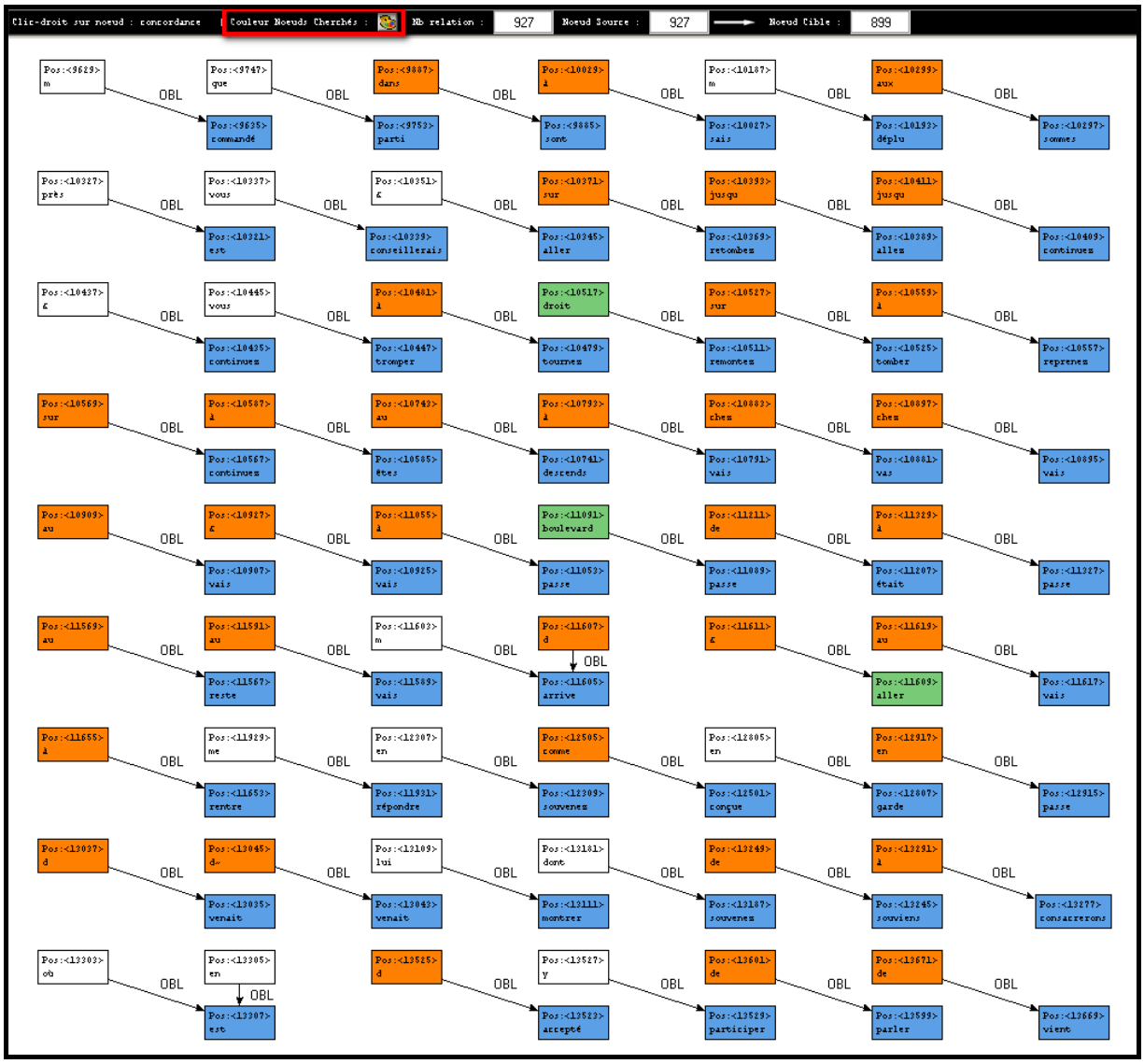


Figure 203 : *Le Trameur* Recherche dans un graphe de relation (3)

25.7 Affichage de relation en contexte

Le ***Trameur*** permet des accès à la ***Trame*** via différents éditeurs (éditeur du ***CADRE***, de la **Carte des sections**, du **concordancier**). Ces éditeurs permettent de donner à voir les relations définies sur les items présents dans l'éditeur.

Dans l'exemple qui suit, la figure présente l'éditeur de la **Carte des sections** contenant l'item (en position 15) dont l'annotation n°1 (forme) est « sapin » (le mode aperçu signale des relations définies sur cet item via le niveau d'annotation n°7)

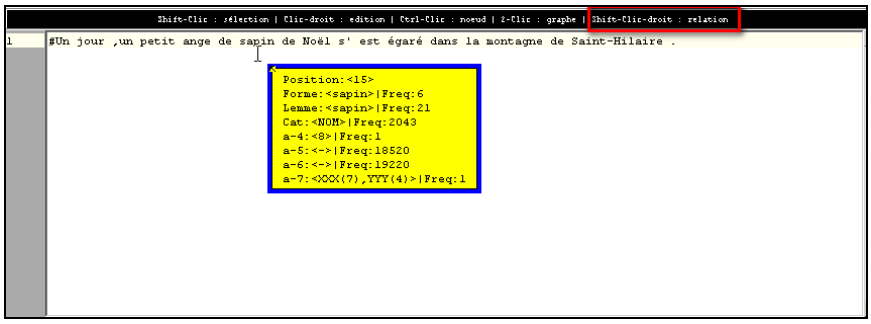


Figure 204 : *Le Trameur* accès aux dépendances via la *Trame*

La combinaison de touches « `shift-clic droit` » active une fenêtre permettant de sélectionner les relations à rechercher sur cet item :

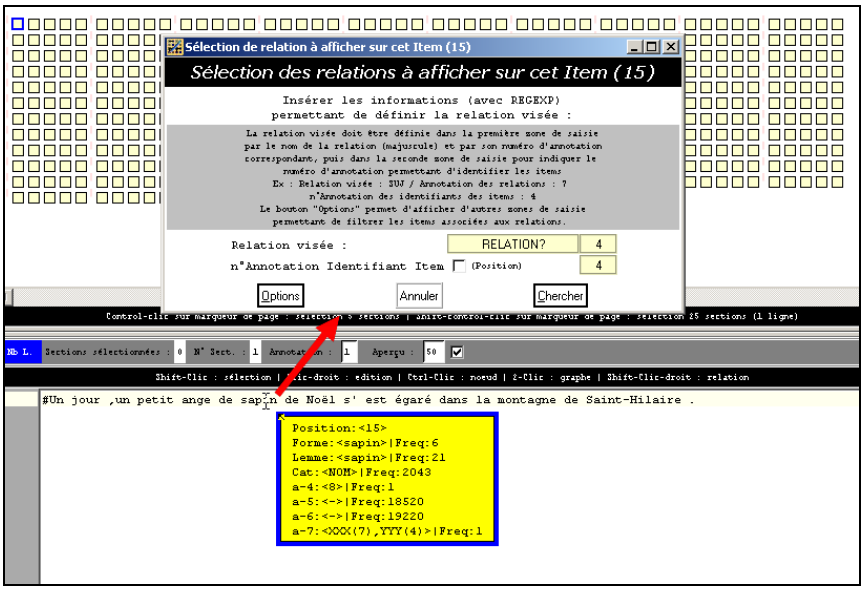


Figure 205 : *Le Trameur* paramétrage de l'affichage des dépendances sur un item de la *Trame* (1)

Comme précédemment, le paramétrage de cette recherche permet de spécifier le nom de la relation visée et le numéro d'annotation dans lequel elle doit être cherchée ; dans la figure qui suit, on cherche les relations XXX ou YYY intégrées dans l'annotation n°7 :

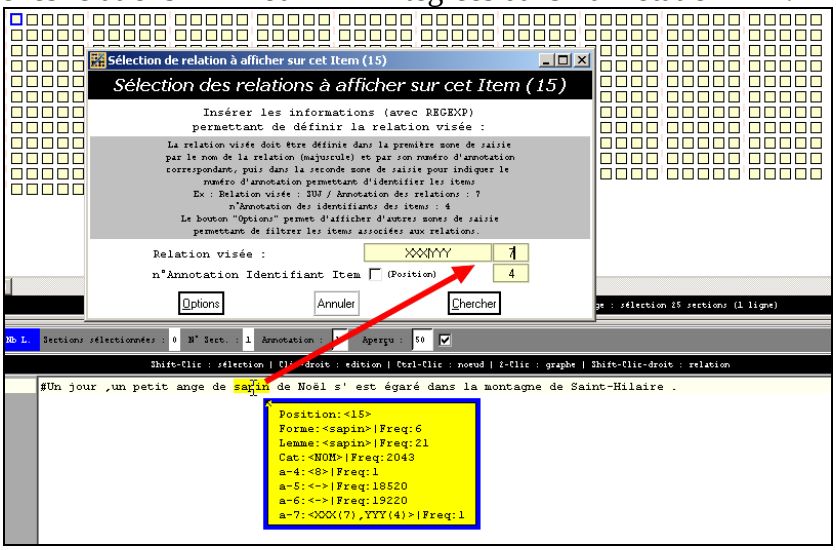


Figure 206 : *Le Trameur* paramétrage de l'affichage des dépendances sur un item de la *Trame* (2)

Le résultat de la recherche produit deux types d'affichage :

- La matérialisation dans l'éditeur des items concernés par relations visées

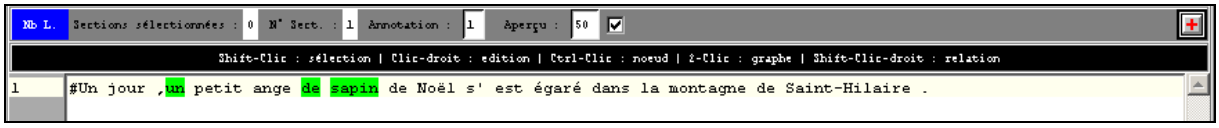

Nb L.
Sections sélectionnées : 0
N° Sect. : 1
Annotation : 1
Aperçu : 50
Shift-Clic : sélection | Clic-droit : edition | Ctrl-Clic : noeud | 2-Clic : graphe | Shift-Clic-droit : relation
1
#Un jour ,un petit ange de sapin de Noël s' est égaré dans la montagne de Saint-Hilaire .

Figure 207 : *Le Trameur* affichage de la dépendance sur la *Trame* (1)

- L'affichage du graphe des relations

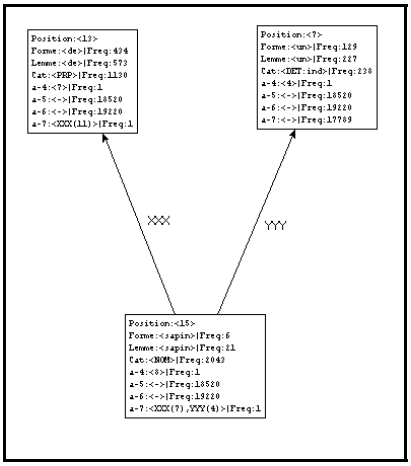


Figure 208 : *Le Trameur* affichage de la dépendance via un graphe (2)

25.8 Recherche de relation en contexte

Les éditeurs de la *Trame* permettent aussi de rechercher d'éventuelles relations sur les items édités.

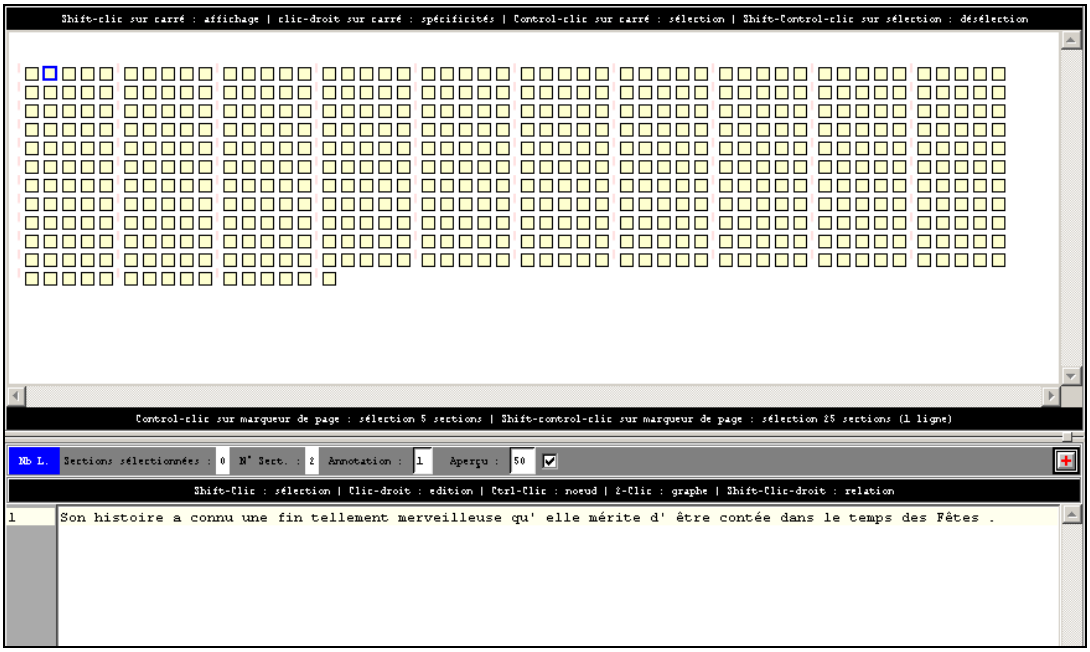

Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : désélection
Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)
Nb L.
Sections sélectionnées : 0
N° Sect. : 2
Annotation : 1
Aperçu : 50
Shift-Clic : sélection | Clic-droit : edition | Ctrl-Clic : noeud | 2-Clic : graphe | Shift-Clic-droit : relation
1
Son histoire a connu une fin tellement merveilleuse qu' elle mérite d' être contée dans le temps des Fêtes .

Figure 209 : *Le Trameur* recherche de dépendance en contexte sur un item de la *Trame*

Dans l'exemple qui suit, l'éditeur de la **Carte des sections** affiche une section contenant l'item n°44 dans l'annotation n°1 (forme) est « connu ». La combinaison de touches « control-clic droit » active une fenêtre permettant de paramétrer la recherche de relations concernant cet item. Le paramétrage de cette recherche de relations consiste à spécifier le numéro de l'annotation dans laquelle la recherche de relations doit s'effectuer (ici le numéro 7).

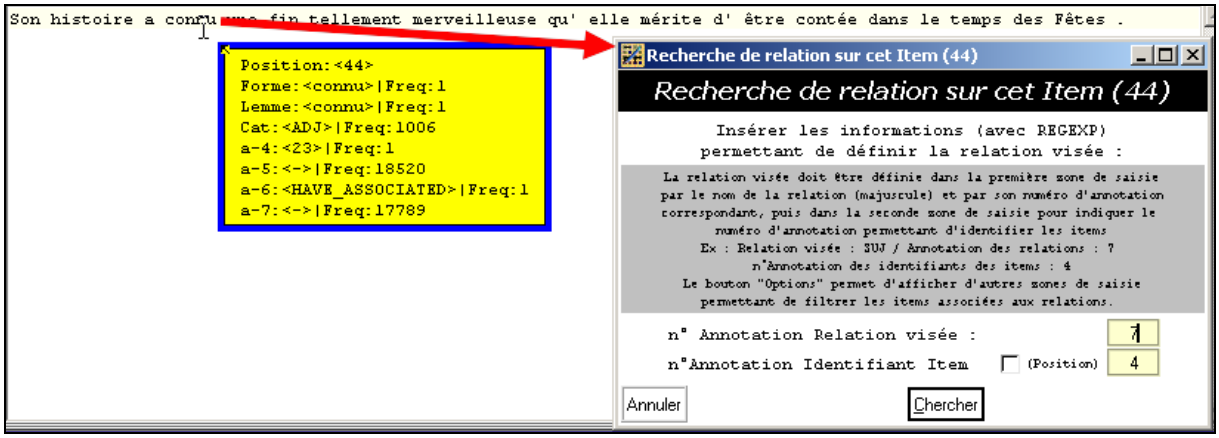

Son histoire a connu une fin tellement merveilleuse qu' elle mérite d' être contée dans le temps des Fêtes .
Position:<44>
Forme:<connu>|Freq:1
Lemme:<connu>|Freq:1
Cat:<ADJ>|Freq:1006
a-4:<23>|Freq:1
a-5:<->|Freq:18520
a-6:<HAVE_ASSOCIATED>|Freq:1
a-7:<->|Freq:17789
Recherche de relation sur cet Item (44)
Recherche de relation sur cet Item (44)
Insérer les informations (avec REGEXP)
permettant de définir la relation visée :
La relation visée doit être définie dans la première zone de saisie par le nom de la relation (majuscule) et par son numéro d'annotation correspondant, puis dans la seconde zone de saisie pour indiquer le numéro d'annotation permettant d'identifier les items
Ex : Relation visée : SUJ / Annotation des relations : 7
n°Annotation des identifiants des items : 4
Le bouton "Options" permet d'afficher d'autres zones de saisie permettant de filtrer les items associées aux relations.
n° Annotation Relation visée : 7
n°Annotation Identifiant Item (Position) 4
Annuler
Chercher

Figure 210 : *Le Trameur* paramétrage de la recherche de dépendance sur un item de la *Trame* (1)

Le résultat de la recherche produit là aussi 2 types de résultats.

1. Affichage du graphe des relations concernant le nœud initialement visé

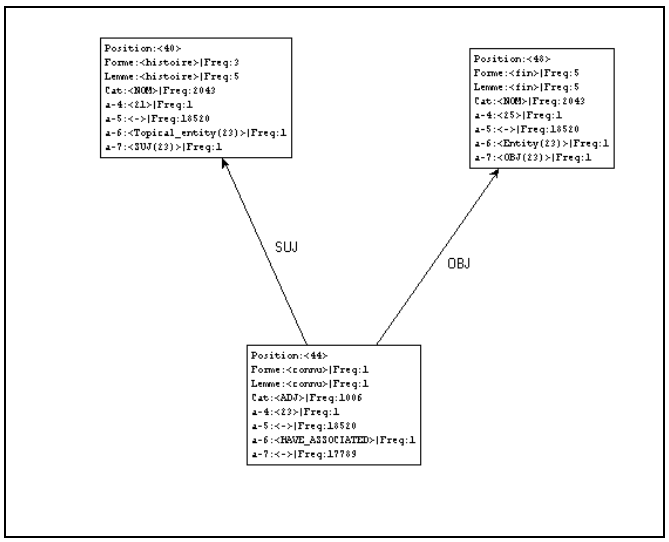


Figure 211 : *Le Trameur* affichage des dépendances via un graphe (1)

2. La matérialisation dans l'éditeur des items concernés

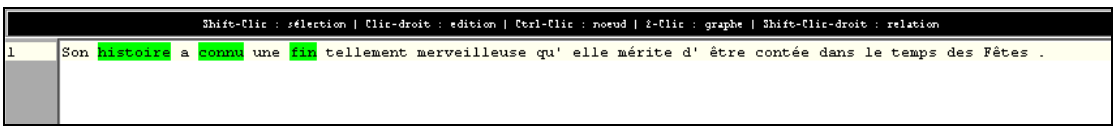

Shift-Clic : sélection | Clic-droit : edition | Ctrl-Clic : noeud | 2-Clic : graphe | Shift-Clic-droit : relation
1
Son histoire a connu une fin tellement merveilleuse qu' elle mérite d' être contée dans le temps des Fêtes .

Figure 212 : *Le Trameur* affichage des dépendances en contexte sur la *Trame*

De même la recherche de dépendance peut aussi être réalisée sur le concordancier. Si on considère la concordance suivante (en haut de la figure) :

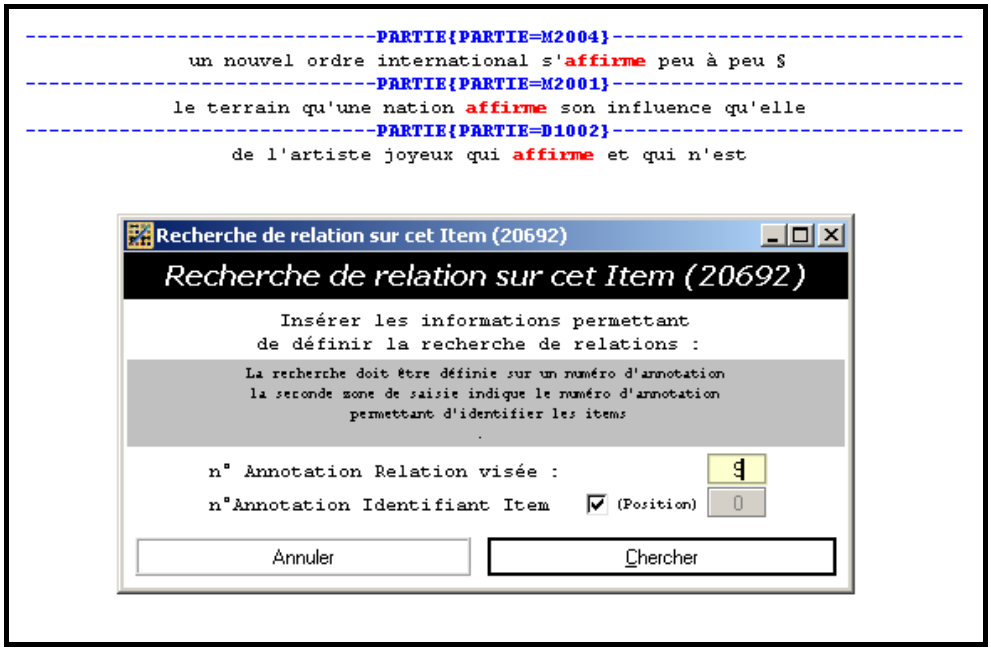

----------PARTIE{PARTIE=M2004}----------
un nouvel ordre international s'affirme peu à peu §
----------PARTIE{PARTIE=M2001}----------
le terrain qu'une nation affirme son influence qu'elle
----------PARTIE{PARTIE=D1002}----------
de l'artiste joyeux qui affirme et qui n'est
Recherche de relation sur cet Item (20692)
Recherche de relation sur cet Item (20692)
Insérer les informations permettant
de définir la recherche de relations :
La recherche doit être définie sur un numéro d'annotation
la seconde zone de saisie indique le numéro d'annotation
permettant d'identifier les items
n° Annotation Relation visée :
9
n°Annotation Identifiant Item
(Position)
0
Annuler
Chercher

Figure 213 : *Le Trameur* paramétrage de la recherche de dépendance sur un item de la *Trame* (2)

La combinaison de touches « `control-clic droit` » sur un item de la concordance (ici l'item visé est la première occurrence de la forme graphique « affirme ») permet de rechercher toutes les relations de dépendance pointant sur cet item. On commence par indiquer où chercher les relations de dépendance (numéro d'annotation portant ce type d'information : ici le n°9) et comment sont indexer les items sur la trame (ici par leur position). La recherche conduit à la production d'un graphe mettant au jour toutes les relations sur l'item visé :

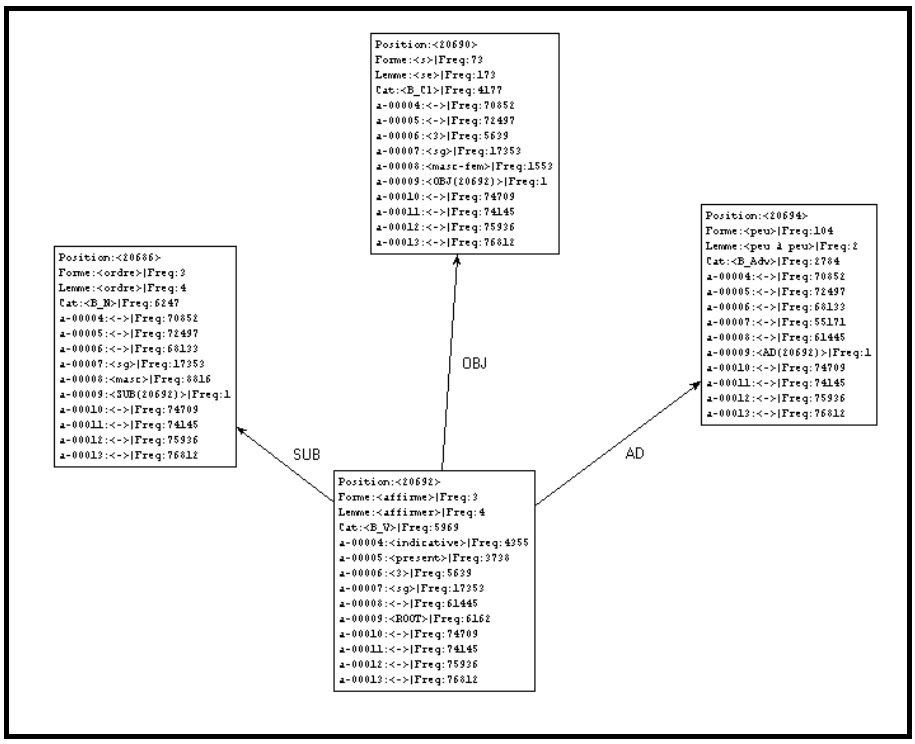


Figure 214 : *Le Trameur* affichage des dépendances via un graphe (2)

Le graphe donne à voir les 3 relations pointant sur l'item.

25.9 Graphe de relations en contexte

Une fonctionnalité permet de donner à voir (sous la forme d'un graphe) l'ensemble des relations disponibles sur une zone de la ***Trame*** éditée dans la **Carte des sections**. La base étant chargée, la figure suivante donne à voir la **Carte des sections** et l'édition d'une section :

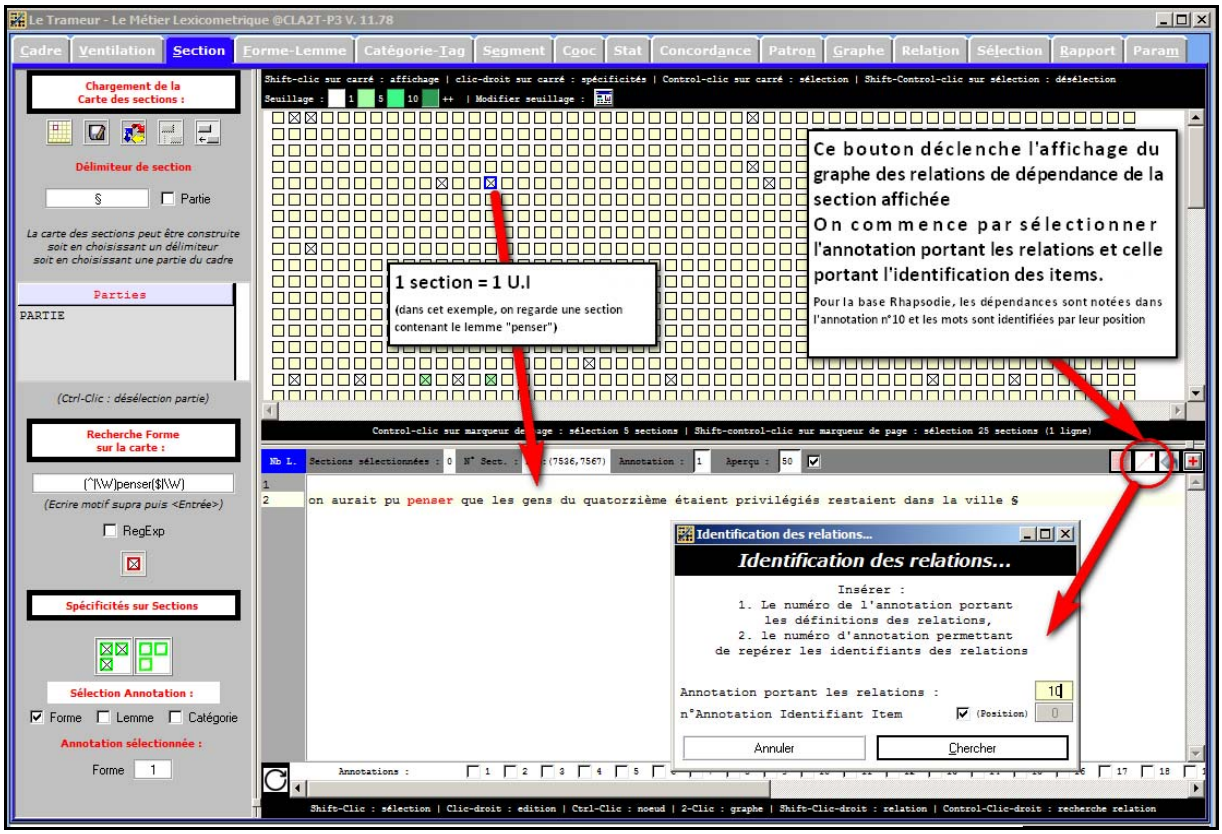

Le Trameur - Le Métier Lexicométrique @CLA2T-P3 V. 11.78
Cadre
Ventilation
Section
Forme-Lemme
Catégorie-Tag
Segment
Cooc
Stat
Concordance
Patron
Graphe
Relation
Sélection
Rapport
Param
Chargement de la Carte des sections :
Délimiteur de section
Partie
La carte des sections peut être construite soit en choisissant un délimiteur soit en choisissant une partie du cadre
Parties
PARTIE
(Ctrl-Clic : désélection partie)
Recherche Forme sur la carte :
(^|\W)penser($|\W)
(Ecrire motif supra puis <Entrée>)
RegExp
Spécificités sur Sections
Sélection Annotation :
Forme
Lemme
Catégorie
Annotation sélectionnée :
Forme
1
1 section = 1 U.I
(dans cet exemple, on regarde une section contenant le lemme "penser")
Ce bouton déclenche l'affichage du graphe des relations de dépendance de la section affichée
On commence par sélectionner l'annotation portant les relations et celle portant l'identification des items.
Pour la base Rhapsodie, les dépendances sont notées dans l'annotation n°10 et les mots sont identifiées par leur position
on aurait pu penser que les gens du quatorzième étaient privilégiés restaient dans la ville §
Identification des relations...
Identification des relations...
Insérer :
1. Le numéro de l'annotation portant les définitions des relations,
2. le numéro d'annotation permettant de repérer les identifiants des relations
Annotation portant les relations :
10
n°Annotation Identifiant Item
(Position)
Annuler
Chercher

Figure 215 : *Le Trameur* affichage d'un graphe de dépendances en contexte (1)

Dès qu'une section est éditée, une icône est activée et permet de déclencher l'affichage du graphe des relations de dépendance entre les items de la zone textuelle éditée.

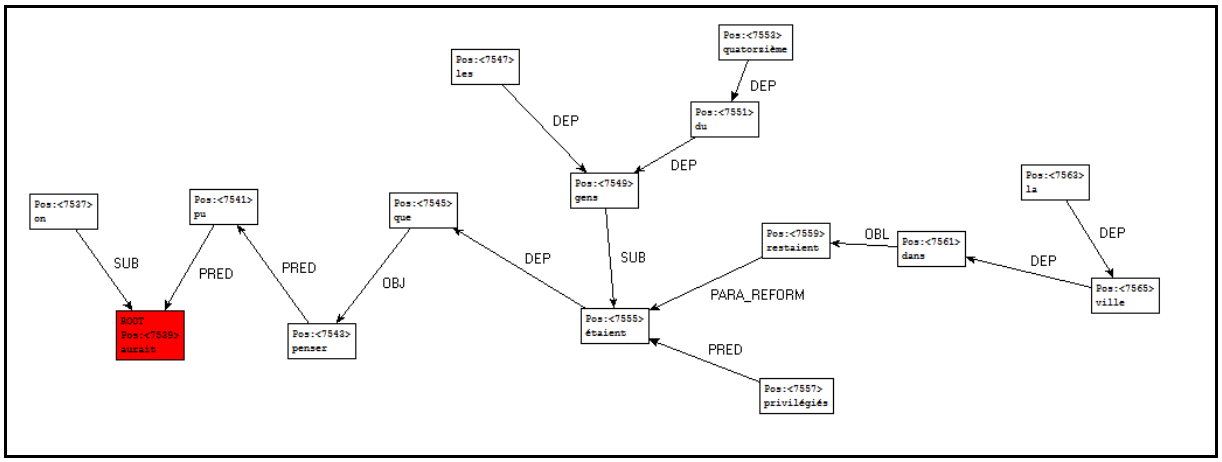


Figure 216 : *Le Trameur* affichage d'un graphe de dépendances en contexte (2)

25.10 Visualisation de toutes les relations dans un contexte donné

Une fonctionnalité complémentaire permet de donner à voir (sous la forme d'une matrice) l'ensemble des relations disponibles sur une zone de la ***Trame*** éditée dans la **Carte des sections**. La base étant chargée, la figure suivante donne à voir la **Carte des sections** (le caractère point est utilisé comme délimiteur de sections) et l'édition de la première section :

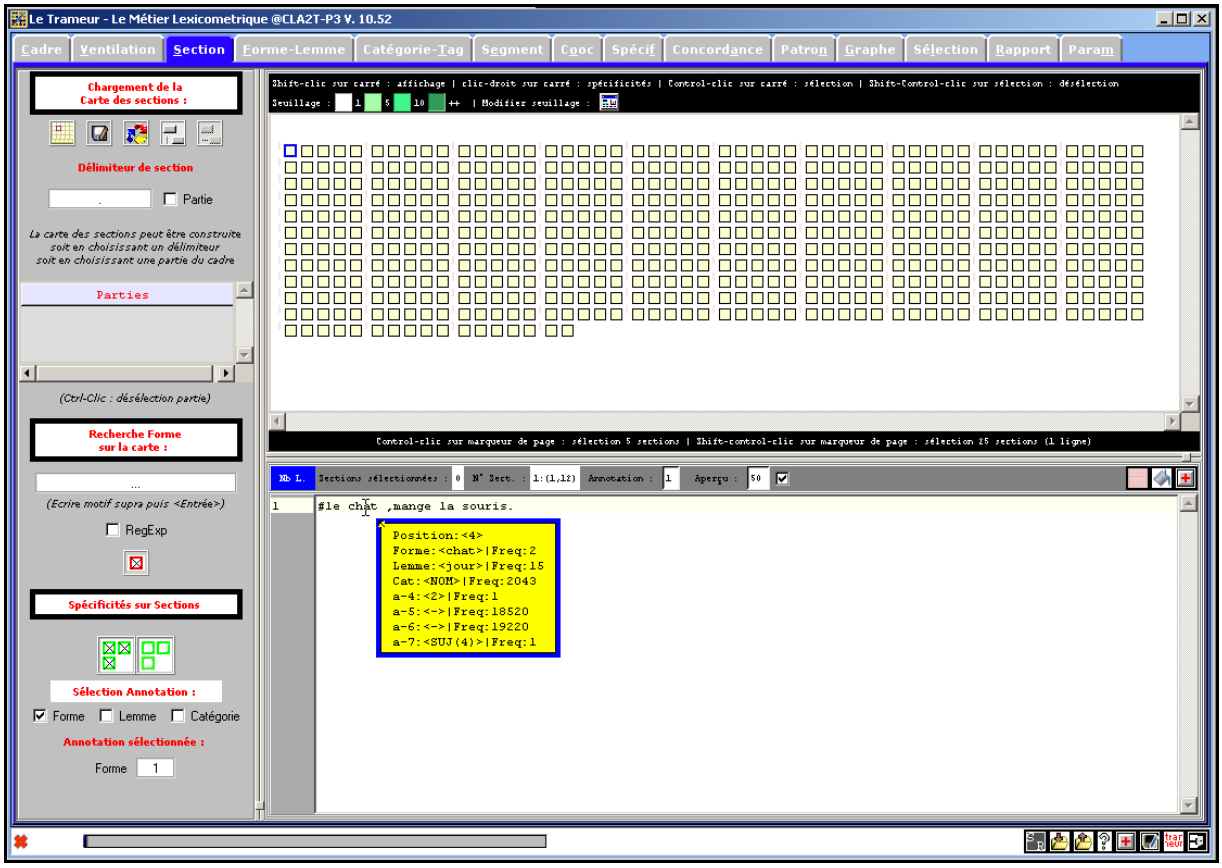


Figure 217 : *Le Trameur* affichage global des dépendances en contexte (1)

Dès qu'une section est éditée, une icône est activée et permet de déclencher l'affichage des relations de dépendance entre les items de la zone textuelle éditée.

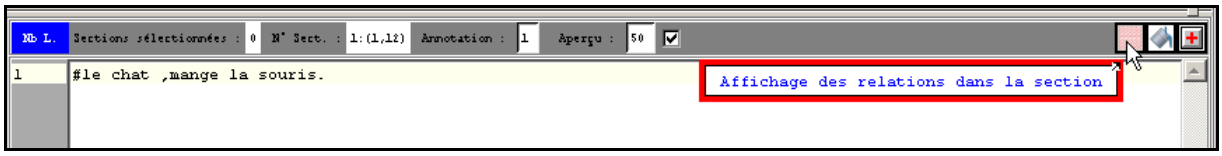


Figure 218 : *Le Trameur* affichage global des dépendances en contexte (2)

Un menu permet de sélectionner le numéro d'annotation portant les relations visées (ici l'annotation n°7) et la manière dans les items sont identifiables dans ces relations (ici via l'annotation n°4)

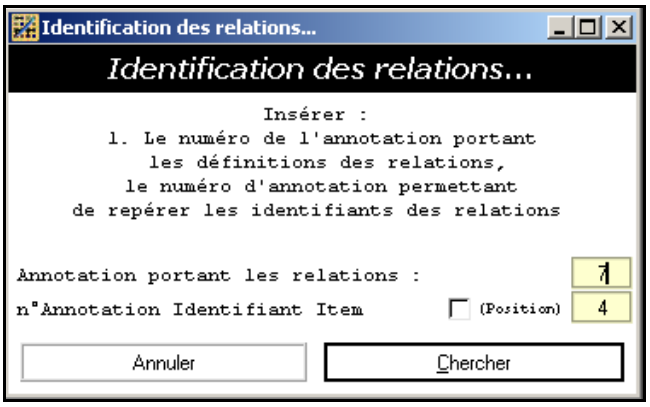


Figure 219 : *Le Trameur* affichage global des dépendances en contexte (3)

L'activation de la recherche produit une matrice donnant à voir les relations disponibles sur la section visée :

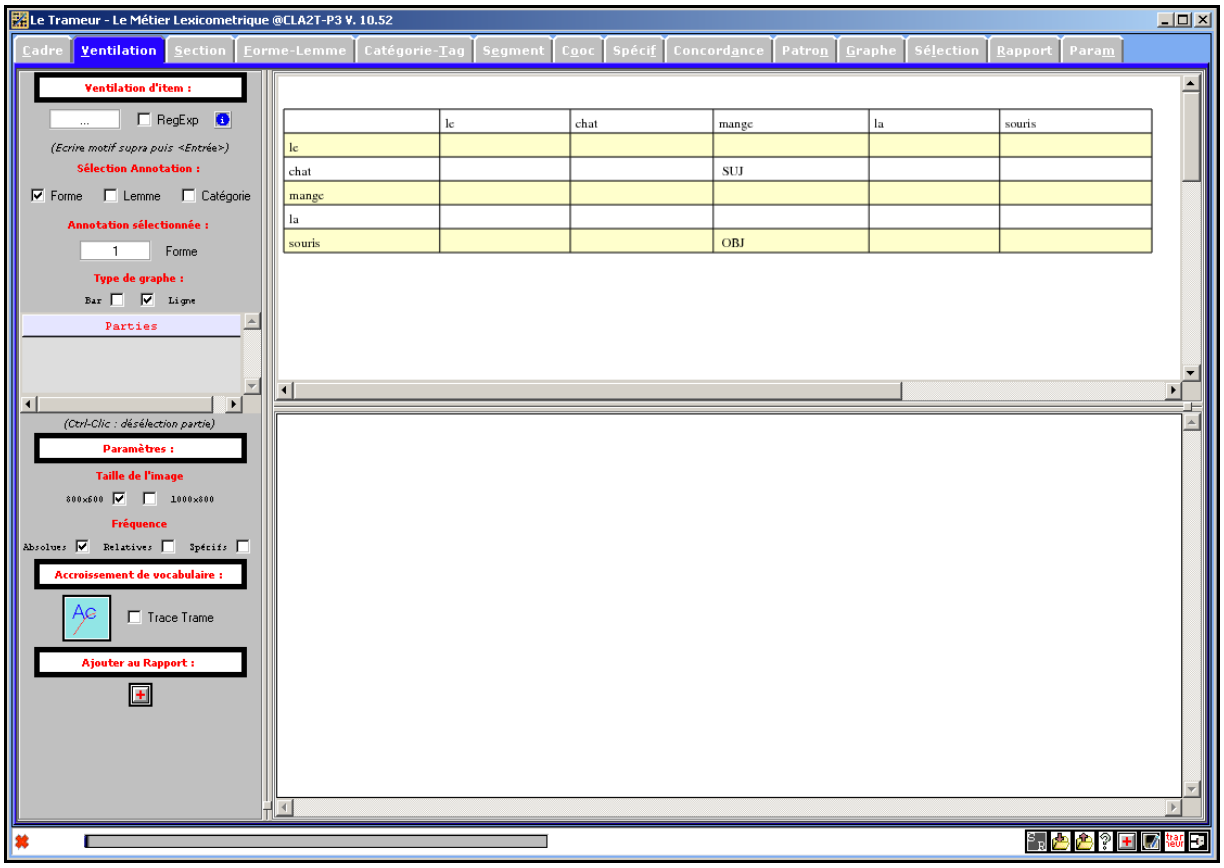


Figure 220 : *Le Trameur* affichage global des dépendances en contexte (4)

25.11 Recherche de collocation : spécificités sur relation

Une version particulière du module de calcul des cooccurrences permet de prendre en compte les relations entre les items de la ***Trame*** :

Etant donné une forme pôle, la recherche de ses cooccurrents peut être contrainte par la prise en compte d'une relation entre ce pôle est ses candidats cooccurrents. Pour un item X donné (le pôle), on s'intéresse aux items Y cooccurrents de X et en relation REL avec X (les collocatifs de X) :

X -> REL -> Y

La figure suivante présente les cooccurrents du lemme penser en contraignant ses cooccurrents à être en position d'objet :

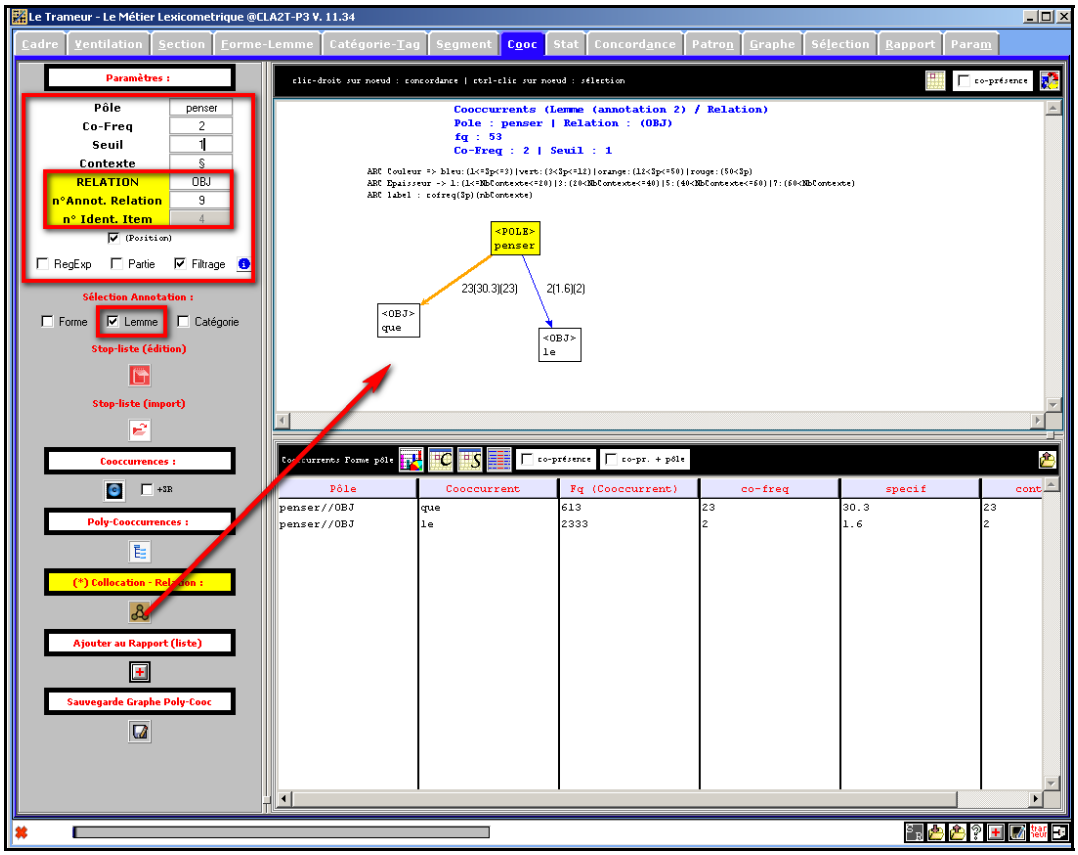


Figure 221 : *Le Trameur* Recherche de collocation (1)

La relation entre le pôle et ses cooccurrents peut être « non déterminée » par avance, dans la figure suivante, la relation est exprimée sous la forme : .* (regexp) ; elle vise donc toutes les relations du type : `penser-OBJ-y`, `penser-SUB-y` etc.

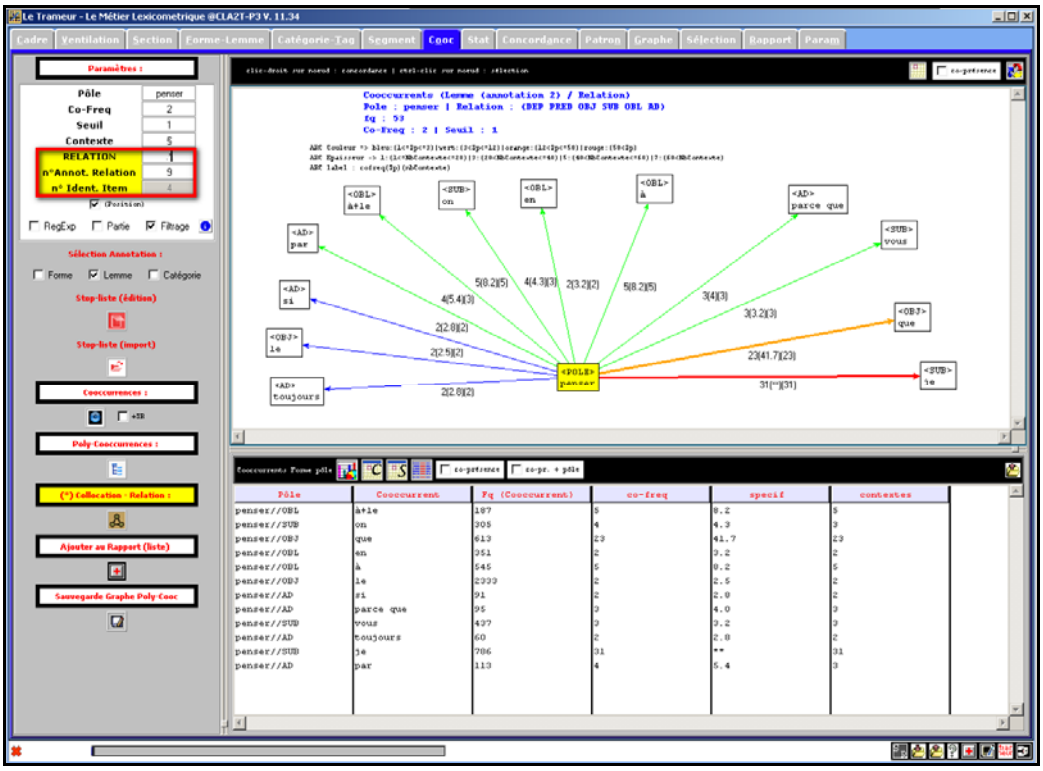


Figure 222 : *Le Trameur* Recherche de collocation (2)

Une fois le graphe construit, un `clic-droit` sur un des noeuds montrent les contextes dans lesquels la relation se réalise (i.e les contextes utilisés par le calcul).

Dans cet exemple, 3 contextes dans lesquels *vous* est sujet de *penser* (*cf* arc `penser -> vous`) ont été mis au jour, un `clic-droit` sur le nœud *vous* les montrent dans le concordancier (les 2 items de la relation y sont colorés automatiquement : *vous* en rouge (pôle de la concordance) et *penser* en vert).

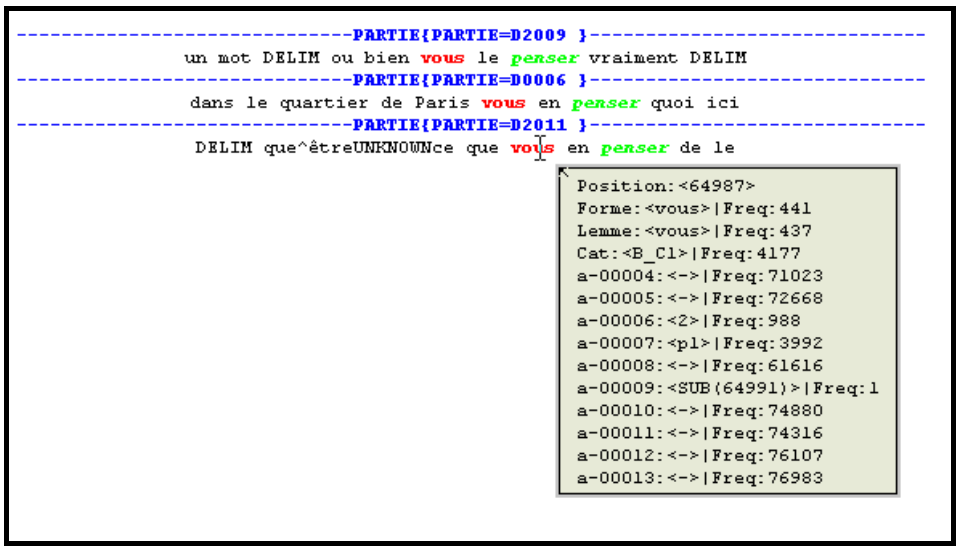


Figure 223 : *Le Trameur* Recherche de collocation (3)

25.12 Recherche source-cible d'une relation

L'onglet **Relation** contient d'autres fonctionnalités permettant de traiter les relations de dépendance. Une première permet de lister les termes *source* et/ou *cible* d'une relation visée. On peut soit chercher la liste des termes *source* de cette relation, soit la liste des termes *cible*.

Dans la figure suivante, on s'intéresse à la relation `OBJ` (les annotations de dépendance étant portées ici par l'annotation n°10 de la base traitée et les items mis en relation sont indexés par leur position sur la ***Trame***), la requête « chercher » produit la liste des items de la ***Trame*** portant cette annotation (*cf* colonne a-00010) :

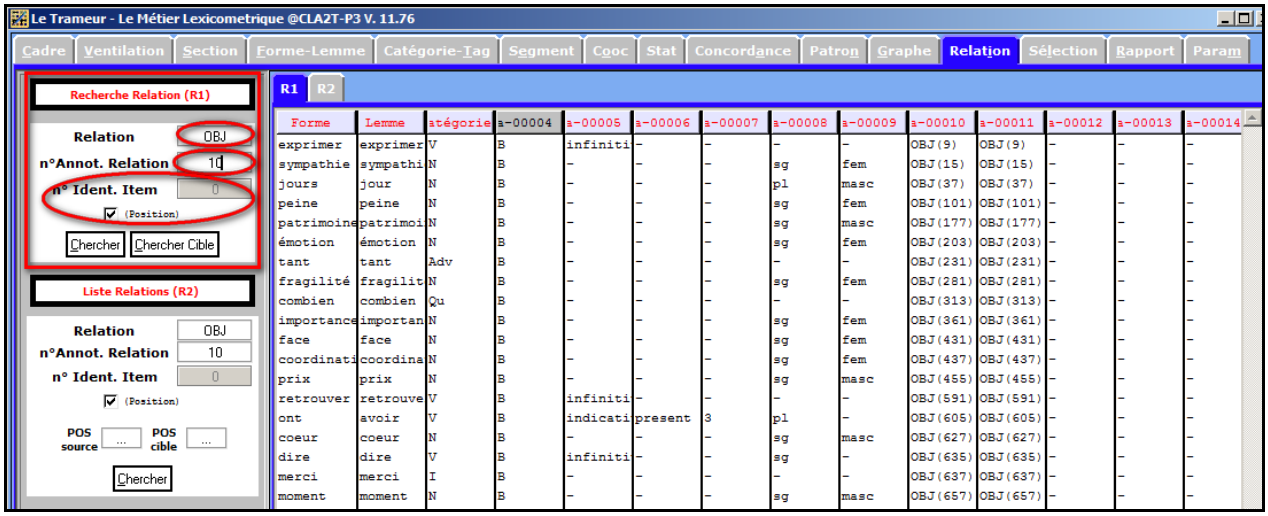

Le Trameur - Le Métier Lexicometrique @CLA2T-P3 V. 11.76
Cadre
Ventilation
Section
Forme-Lemme
Catégorie-Tag
Segment
Cooc
Stat
Concordance
Patron
Graphe
Relation
Sélection
Rapport
Param
Recherche Relation (R1)
Relation
OBJ
n°Annot. Relation
10
n° Ident. Item
0
(Position)
Chercher
Chercher Cible
Liste Relations (R2)
Relation
OBJ
n°Annot. Relation
10
n° Ident. Item
0
(Position)
POS source
POS cible
Chercher
R1
R2
Forme
Lemme
atégorie
a-00004
a-00005
a-00006
a-00007
a-00008
a-00009
a-00010
a-00011
a-00012
a-00013
a-00014
exprimer exprimer V B infiniti - - - - OBJ(9) OBJ(9) - - -
sympathie sympathi N B - - - sg fem OBJ(15) OBJ(15) - - -
jours jour N B - - - pl masc OBJ(37) OBJ(37) - - -
peine peine N B - - - sg fem OBJ(101) OBJ(101) - - -
patrimoine patrimoi N B - - - sg masc OBJ(177) OBJ(177) - - -
émotion émotion N B - - - sg fem OBJ(203) OBJ(203) - - -
tant tant Adv B - - - - - OBJ(231) OBJ(231) - - -
fragilité fragilit N B - - - sg fem OBJ(281) OBJ(281) - - -
combien combien Qu B - - - - - OBJ(313) OBJ(313) - - -
importance importan N B - - - sg fem OBJ(361) OBJ(361) - - -
face face N B - - - sg fem OBJ(431) OBJ(431) - - -
coordinati coordina N B - - - sg fem OBJ(437) OBJ(437) - - -
prix prix N B - - - sg masc OBJ(455) OBJ(455) - - -
retrouver retrouve V B infiniti - - - - OBJ(591) OBJ(591) - - -
ont avoir V B indicati present 3 pl - OBJ(605) OBJ(605) - - -
coeur coeur N B - - - sg masc OBJ(627) OBJ(627) - - -
dire dire V B infiniti - - sg - OBJ(635) OBJ(635) - - -
merci merci I B - - - - - OBJ(637) OBJ(637) - - -
moment moment N B - - - sg masc OBJ(657) OBJ(657) - - -

Figure 224 : *Le Trameur* liste de relations

Dans la figure suivante, on s'intéresse à la même relation, la requête « chercher cible » produit cette fois-ci les cibles des relations portées par les items précédents :

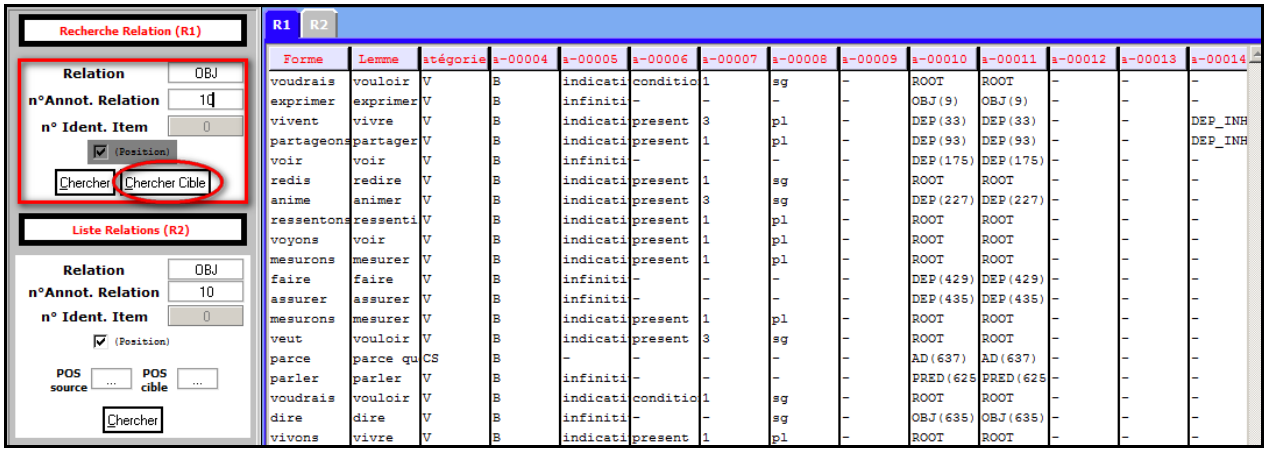

Recherche Relation (R1)
Relation
OBJ
n°Annot. Relation
10
n° Ident. Item
0
(Position)
Chercher
Chercher Cible
Liste Relations (R2)
Relation
OBJ
n°Annot. Relation
10
n° Ident. Item
0
(Position)
POS source
POS cible
Chercher
R1
R2
Forme
Lemme
atégorie
a-00004
a-00005
a-00006
a-00007
a-00008
a-00009
a-00010
a-00011
a-00012
a-00013
a-00014
voudrais vouloir V B indicati conditio 1 sg - ROOT ROOT - - -
exprimer exprimer V B infiniti - - - - OBJ(9) OBJ(9) - - -
vivent vivre V B indicati present 3 pl - DEP(33) DEP(33) - - DEP_INH
partageons partager V B indicati present 1 pl - DEP(93) DEP(93) - - DEP_INH
voir voir V B infiniti - - - - DEP(175) DEP(175) - - -
redis redire V B indicati present 1 sg - ROOT ROOT - - -
anime animer V B indicati present 3 sg - DEP(227) DEP(227) - - -
ressentons ressenti V B indicati present 1 pl - ROOT ROOT - - -
voyons voir V B indicati present 1 pl - ROOT ROOT - - -
mesurons mesurer V B indicati present 1 pl - ROOT ROOT - - -
faire faire V B infiniti - - - - DEP(429) DEP(429) - - -
assurer assurer V B infiniti - - - - DEP(435) DEP(435) - - -
mesurons mesurer V B indicati present 1 pl - ROOT ROOT - - -
veut vouloir V B indicati present 3 sg - ROOT ROOT - - -
parce parce qu CS B - - - - - AD(637) AD(637) - - -
parler parler V B infiniti - - - - PRED(625 PRED(625 - - -
voudrais vouloir V B indicati conditio 1 sg - ROOT ROOT - - -
dire dire V B infiniti - - sg - OBJ(635) OBJ(635) - - -
vivons vivre V B indicati present 1 pl - ROOT ROOT - - -

Figure 225 : *Le Trameur* liste de relations (cible)

On peut enfin construire la liste des termes de `POS` contraints par un relation de dépendance. Dans la figure qui suit, la relation visée est la relation `OBJ` ; on paramètre ensuite les catégories *source* et *cible* visées (annotation n°3) visées (ici `.*` permet de les récupérer toutes).

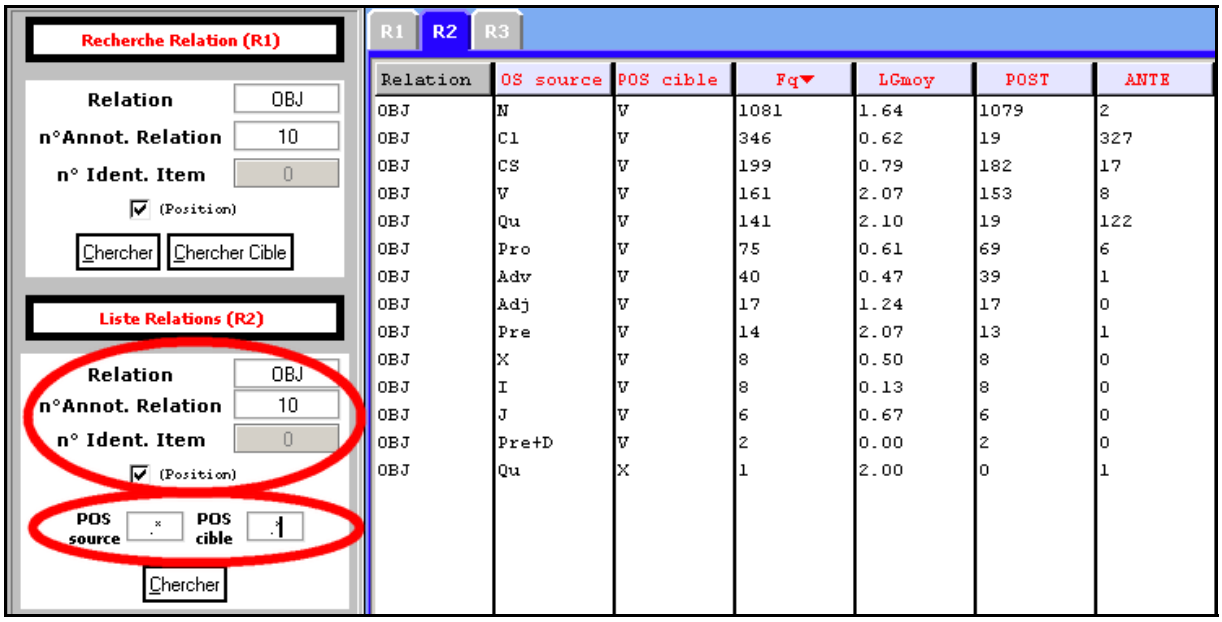

Recherche Relation (R1)
Relation
OBJ
n°Annot. Relation
10
n° Ident. Item
0
(Position)
Chercher
Chercher Cible
Liste Relations (R2)
Relation
OBJ
n°Annot. Relation
10
n° Ident. Item
0
(Position)
POS source
POS cible
Chercher
R1
R2
R3
Relation | OS source | POS cible | Fq | LGmoy | POST | ANTE
OBJ | N | V | 1081 | 1.64 | 1079 | 2
OBJ | Cl | V | 346 | 0.62 | 19 | 327
OBJ | CS | V | 199 | 0.79 | 182 | 17
OBJ | V | V | 161 | 2.07 | 153 | 8
OBJ | Qu | V | 141 | 2.10 | 19 | 122
OBJ | Pro | V | 75 | 0.61 | 69 | 6
OBJ | Adv | V | 40 | 0.47 | 39 | 1
OBJ | Adj | V | 17 | 1.24 | 17 | 0
OBJ | Pre | V | 14 | 2.07 | 13 | 1
OBJ | X | V | 8 | 0.50 | 8 | 0
OBJ | I | V | 8 | 0.13 | 8 | 0
OBJ | J | V | 6 | 0.67 | 6 | 0
OBJ | Pre+D | V | 2 | 0.00 | 2 | 0
OBJ | Qu | X | 1 | 2.00 | 0 | 1

Figure 226 : *Le Trameur* liste de POS en relation

Le résultat produit donne à voir les fréquences des patrons de dépendance (via leur POS) : par exemple ici, la relation de dépendance OBJ met 40 fois en relation un item de type Adv et un item de type V avec : Adv–OBJ->V ; dans la base traitée ici, la relation OBJ met majoritairement en relation les items de type V et les items de type N : N–OBJ->V (fq : 1081).

Les trois dernières colonnes donnent des informations complémentaires sur chaque type de relation :

- **LGmoy** (longueur moyenne) : le nombre de mots entre l'item source (dépendant) et l'item cible (gouverneur).
- **POST** : le nombre de relation avec un dépendant postposé.
- **ANTE** : le nombre de relation avec un dépendant antéposé.

Enfin, dans la figure suivante, la requête exprimée permet d'afficher, pour un item donné (ici le lemme « penser »), les dépendants de cet item dans la relation de dépendance visée (ici SUB ou OBJ) :

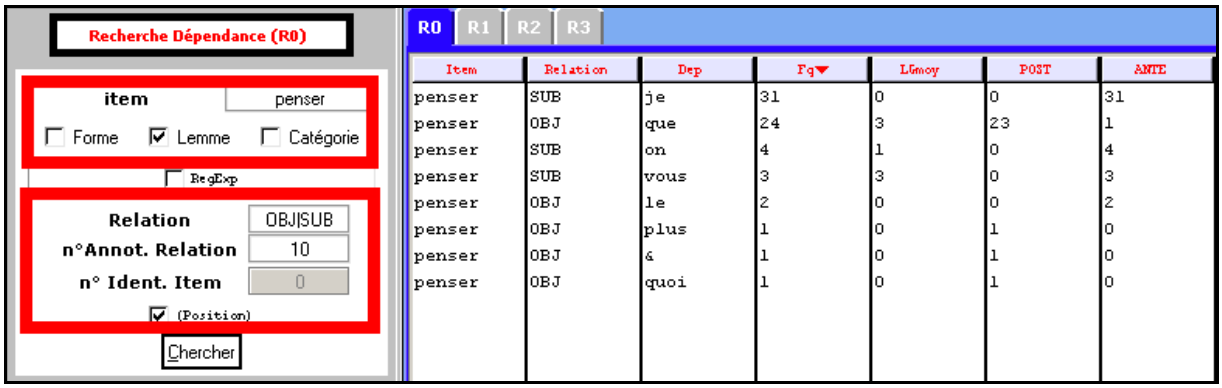

Recherche Dépendance (R0)
item
penser
Forme
Lemme
Catégorie
RegExp
Relation
OBJ|SUB
n°Annot. Relation
10
n° Ident. Item
0
(Position)
Chercher
R0
R1
R2
R3
Item | Relation | Dep | Fq | LGmoy | POST | ANTE
penser | SUB | je | 31 | 0 | 0 | 31
penser | OBJ | que | 24 | 3 | 23 | 1
penser | SUB | on | 4 | 1 | 0 | 4
penser | SUB | vous | 3 | 3 | 0 | 3
penser | OBJ | le | 2 | 0 | 0 | 2
penser | OBJ | plus | 1 | 0 | 1 | 0
penser | OBJ | à | 1 | 0 | 1 | 0
penser | OBJ | quoi | 1 | 0 | 1 | 0

Figure 227 : *Le Trameur* les dépendants d'un item (forme, lemme ou catégorie)

Pour produire la liste des dépendants de plusieurs items, il est possible de les décrire *via* une expression régulière (ici tous les mots commençant par « pens ») :

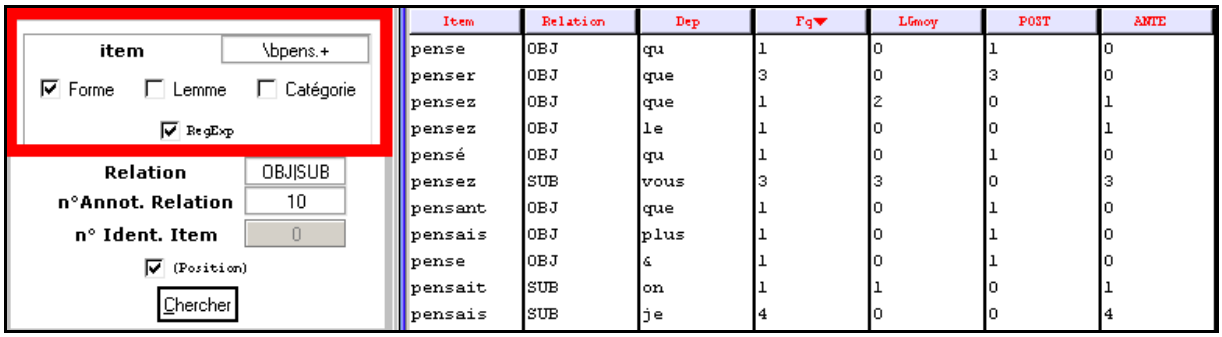


Figure 228 : *Le Trameur* les dépendants d'une liste d'items (forme, lemme ou catégorie)

Les items de la liste peuvent être projetés sur le concordancier via le raccourci clavier « shift-clic » sur le ou les items visé(s) :

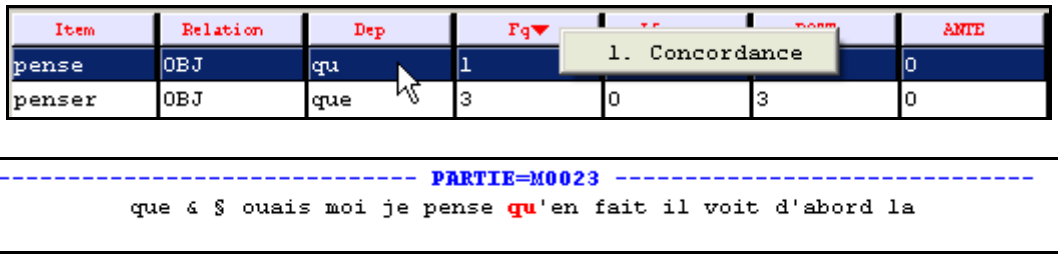

Item Relation Dep Fq ANTE
pense OBJ qu 1 0
penser OBJ que 3 0 3 0
1. Concordance
---------- PARTIE=M0023 ----------
que & § ouais moi je pense qu'en fait il voit d'abord la

Figure 229 : *Le Trameur* retour en contexte (les dépendants d'un item)

26 Le *Cadre* : visualisation, réorganisation

Le ***Cadre*** est présenté supra comme une interface permettant de définir des accès à la ***Trame***. Nous présentons ici un chantier qui étend les outils et fonctionnalités associés à la ***Trame***. Il s'agit essentiellement de permettre à l'utilisateur de réorganiser le ***Cadre*** afin de conduire des opérations supplémentaires à partir de la réorganisation réalisée en agissant dynamiquement sur la structuration du ***Cadre***. Nous illustrerons ces modifications sur 2 fonctionnalités déjà disponibles :

- Réorganisation dynamique du ***Cadre*** pour construire un nouveau corpus de travail
- Réorganisation dynamique (et provisoire) du ***Cadre*** pour un calcul lexicométrique donné (accroissement du vocabulaire)

L'onglet **CADRE** est composé de 2 volets (2 sous-onglets) : l'un dédié à sa visualisation, l'autre à sa réorganisation.

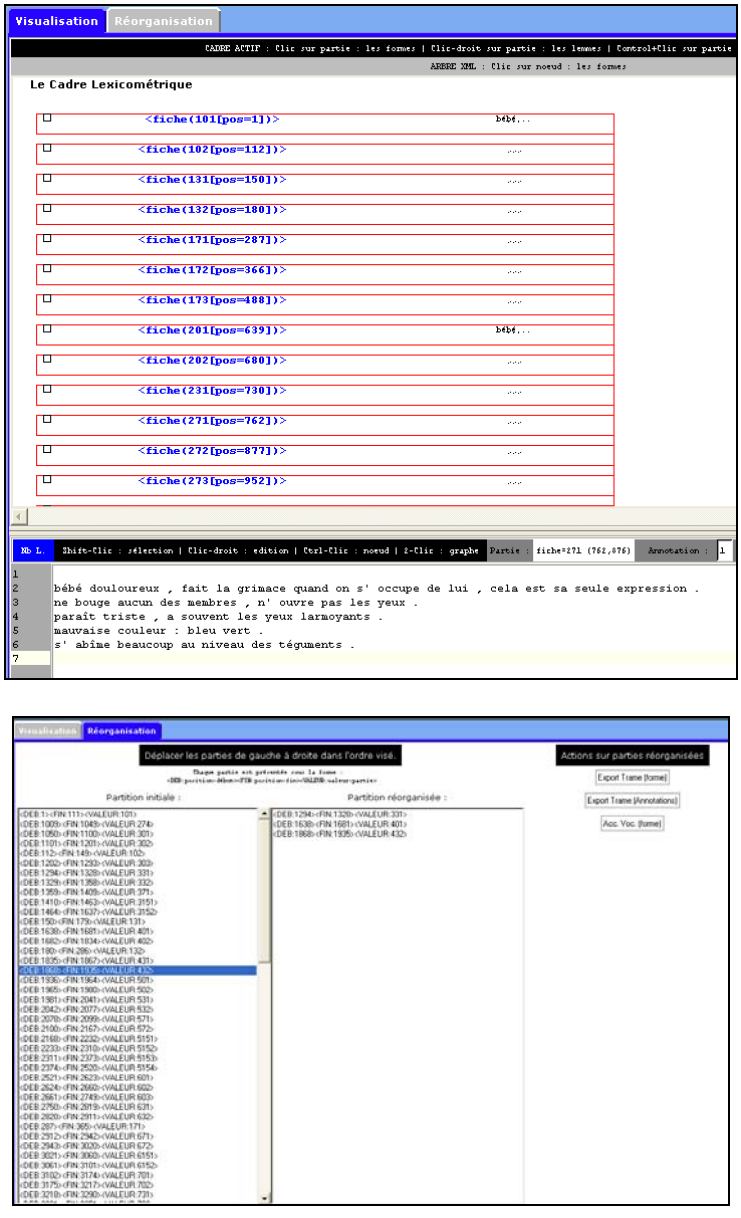

Visualisation
Réorganisation
Le Cadre Lexicométrique
bébé bébé
bébé bébé
1
2 bébé douloureux , fait la grimace quand on s' occupe de lui , cela est sa seule expression .
3 ne bouge aucun des membres , n' ouvre pas les yeux .
4 paraît triste , a souvent les yeux larmoyants .
5 mauvaise couleur : bleu vert .
6 s' abîme beaucoup au niveau des téguments .
7
Déplacer les parties de gauche à droite dans l'ordre visé.
Actions sur parties réorganisées
Partition initiale :
Partition réorganisée :

Figure 230 : *Le Trameur* le *Cadre* (visualisation, réorganisation)

Ce second volet permet de gérer la réorganisation dynamique du ***Cadre***. Une fois la base textométrique chargée, un bouton permet de charger une partition donnée en vue de sa réorganisation dynamique :

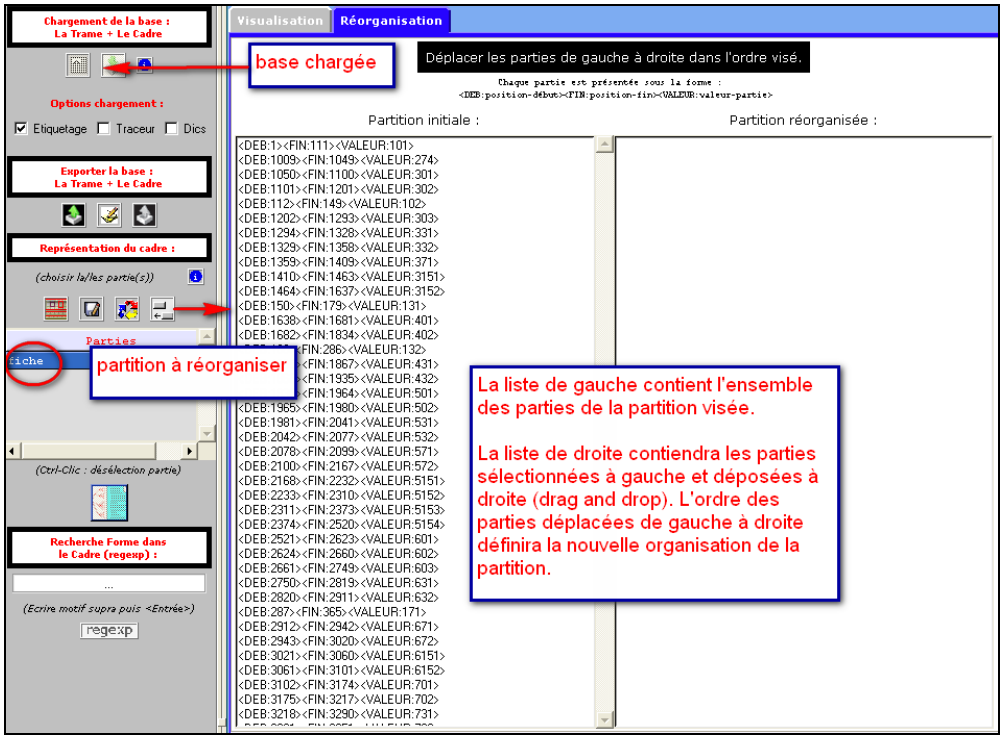

Chargement de la base : La Trame + Le Cadre
Visualisation
Réorganisation
base chargée
Déplacer les parties de gauche à droite dans l'ordre visé.
Options chargement :
Etiquetage
Traceur
Dics
Exporter la base : La Trame + Le Cadre
Représentation du cadre :
(choisir la/les partie(s))
Parties
partition à réorganiser
(Ctrl-Clic : désélection partie)
Recherche Forme dans le Cadre (regexp) :
(Ecrire motif supra puis <Entrée>)
regexp
Partition initiale :
Partition réorganisée :
La liste de gauche contient l'ensemble des parties de la partition visée.
La liste de droite contiendra les parties sélectionnées à gauche et déposées à droite (drag and drop). L'ordre des parties déplacées de gauche à droite définira la nouvelle organisation de la partition.

Figure 231 : *Le Trameur* Chargement d'une partition à réorganiser

L'utilisateur peut donc définir dynamiquement une nouvelle organisation d'une partition initialement définie :

- En modifiant l'ordre des parties
- En ne sélectionnant qu'une sélection restreinte de parties (ici la réorganisation peut être vue comme une opération de *sélection*)

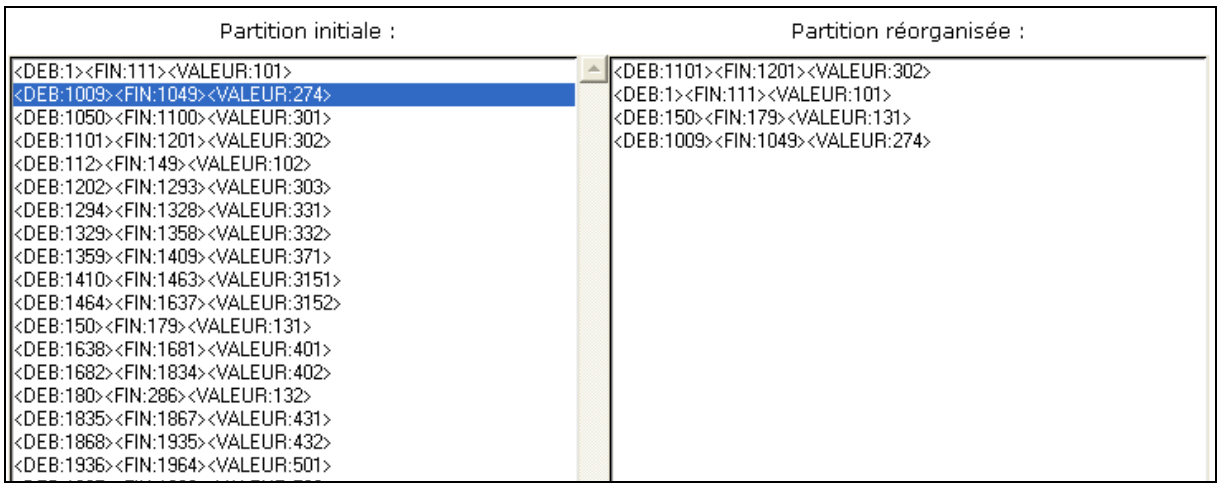


Figure 232 : *Le Trameur* Réorganisation d'une partition

Une fois la réorganisation réalisée, il est possible d'activer différentes opérations sur le **CADRE** réorganisé dynamiquement. Il est important de souligner qu'à ce stage aucune modification n'est faite sur la base initiale : la ***Trame*** et le ***Cadre*** de départ restent dans leur état initial. La réorganisation visée est simplement décrite par la position des parties pour la réorganisation escomptée.

A ce stade du développement, les actions disponibles sont :

1. Export de la base textométrique en tenant compte de la réorganisation visée : cet export permet de regénérer une ***Trame*** et un ***Cadre*** sur la base de l'ordre décrit pour la partition choisie.
   Application : réorganisation d'un corpus pour rétablir par exemple une chronologie textuelle non réalisée dans la phase initiale. Cette réorganisation est vue ici comme un outil de préparation des données (ou de reformatage) avant analyse. Une fois la base exportée, il suffit de recharger la nouvelle base et de poursuivre son exploration.
2. Accroissement du vocabulaire sur la base en tenant compte de la réorganisation réalisée.

Ici, la réorganisation reste virtuelle, elle ne concerne que le calcul opéré. La base initiale reste inchangée. On peut par exemple comparer 2 courbes d'accroissement du vocabulaire sur une même ***Trame*** d'items organisés différemment sur la base de 2 partitions équivalentes mais structurées séquentiellement de manière différente. Le résultat produit donne d'ailleurs à voir le calcul de l'accroissement du vocabulaire sur la base initiale et sur la base réorganisée.

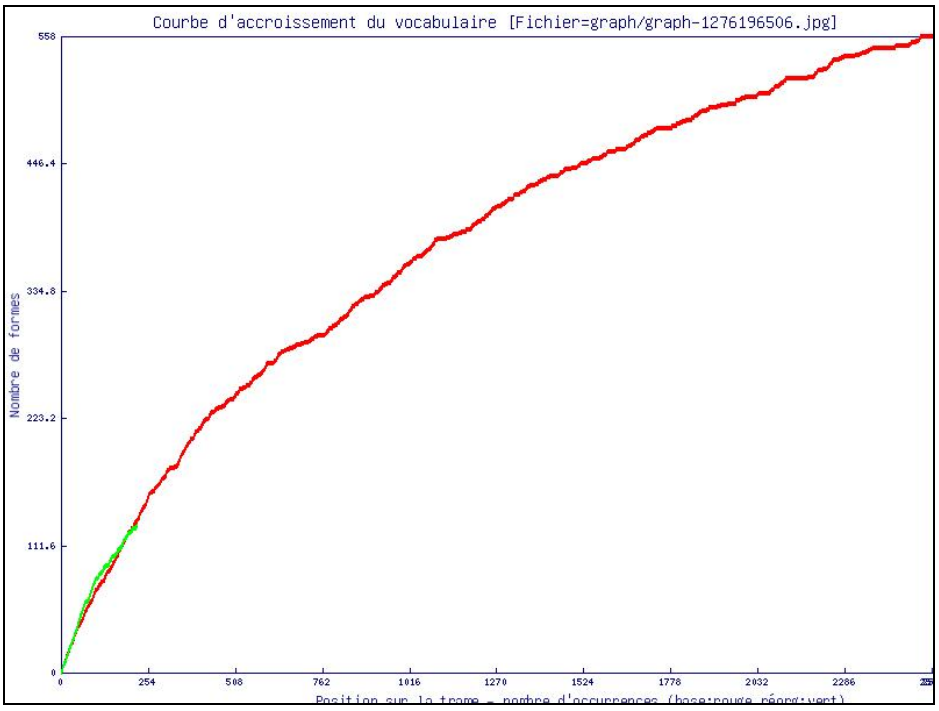


Figure 233 : *Le Trameur* Courbe d'accroissement sur une partition initiale et la même partition tronquée et réorganisée

Dans les 2 cas, l'opération réalisée peut se réaliser sur l'ensemble des parties réorganisées ou sur un sous-ensemble ne couvrant pas totalement l'ensemble de la ***Trame*** initiale, dans ce dernier cas, on extrait un sous corpus ou on calcule l'accroissement du vocabulaire sur une partir de la ***Trame*** initiale réorganisée virtuellement.

27 Exporter une base textométrique

Le ***Trameur*** permet d'importer une *base textométrique i.e.* une ressource textuelle décrivant une ***Trame*** et un ***Cadre*** textométriques. Cette ressource peut être construite par le ***Trameur*** pour construire, à un moment donné, un état textométrique donnant à voir simultanément la ***Trame*** et le ***Cadre*** du texte traité. Elle peut ensuite être réimportée par le processus mis en œuvre pour charger ce type de ressource.

Le processus d'exportation d'une base textométrique est disponible dans l'onglet **CADRE**. La ressource construite après activation de cette exportation est disponible dans le dossier `export` du répertoire contenant le fichier chargé.

On donne à voir ci-dessous le texte de travail « Le dormeur du Val » (chargé avec l'option étiquetage) et le fichier d'exportation obtenu :

Texte initial

```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
<LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
<LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
<LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
<LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut.
<LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
<LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
<LIGNE="VERS11">Nature, berce-le chaudement : il a froid.
<LIGNE="VERS12">Les parfums ne font pas frissonner sa narine ;
<LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
<LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit.
<STRUCTURE="AUTEUR">Arthur Rimbaud
```

Fichier exporté (et ré-importable comme base textométrique) :

```
<?xml version="1.0" encoding="utf8"?>
<baselexicometrique>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base textometrique (le metier = le Cadre et la Trame),
importable et echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
<p>Samedi 02 Juin 2012
 22:15:57
        Ce document n'est pas encore publie.</p>
</publicationStmt>
<sourceDesc>
<title>
<br>Le Trameur 10.38.2. Samedi 02 Juin 2012
 22:15:57 </br></title>
<content><l><br>Fichier traite </br><br> dormeurL3.txt</br></l>
<l><br>Encodage </br><br> iso-8859-1</br></l>
<l><br>Nombre d'items </br><br> 295</br></l>
```

```
<l><br>Nombre de délimiteurs </br><br> 168</br></l>
<l><br>Nombre d'occurrences de forme </br><br> 127</br></l>
<l><br>Nombre de formes </br><br> 97</br></l>
<l><br>Nombre d'hapax </br><br> 80</br></l>
<l><br>Fréquence maximale </br><br> 5</br></l>
<l><br>Forme maximale </br><br> dans</br></l>
<l><br>Délimiteurs </br><br><![CDATA[. ,:;!?/_-"'()[]{}§$|*><=+
«»     ]]></br></l>
<l><br>Etiquetage Treetagger </br><br>OUI</br></l>
<l><br>Langue pour Treetagger </br><br>français</br></l>
</content></sourceDesc>
</fileDesc>
</teiheader>
<Trame>
<codage>utf8</codage>
<delimiteur><![CDATA[. ,:;!?/_-"'()[]{}§$|*><=+
«»     ]]></delimiteur>
<items>
<item type="delim" pos="1"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="2"><f>Le</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="3"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="4"><f>dormeur</f><c>NOM</c><l>dormeur</l></item>
<item type="delim" pos="5"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="6"><f>du</f><c>PRP_det</c><l>du</l></item>
<item type="delim" pos="7"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="8"><f>val</f><c>NOM</c><l>val</l></item>
<item type="delim" pos="9"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="delim" pos="10"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="11"><f>C</f><c>NUM</c><l>C</l></item>
<item type="delim" pos="12"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="delim" pos="13"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="14"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="delim" pos="15"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="16"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="delim" pos="17"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="18"><f>trou</f><c>NOM</c><l>trou</l></item>
<item type="delim" pos="19"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="20"><f>de</f><c>PRP</c><l>de</l></item>
<item type="delim" pos="21"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="22"><f>verdure</f><c>NOM</c><l>verdure</l></item>
<item type="delim" pos="23"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="24"><f>où</f><c>PRO_REL</c><l>où</l></item>
<item type="delim" pos="25"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="26"><f>chante</f><c>VER_pres</c><l>chanter</l></item>
<item type="delim" pos="27"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="28"><f>une</f><c>DET_ART</c><l>un</l></item>
<item type="delim" pos="29"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="30"><f>rivière</f><c>NOM</c><l>rivière</l></item>
<item type="delim" pos="31"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="32"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="delim" pos="33"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="34"><f>Accrochant</f><c>NAM</c><l>Accrochant</l></item>
<item type="delim" pos="35"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="36"><f>follement</f><c>ADV</c><l>follement</l></item>
<item type="delim" pos="37"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="38"><f>aux</f><c>PRP_det</c><l>au</l></item>
<item type="delim" pos="39"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="40"><f>herbes</f><c>VER_pres</c><l>herber</l></item>
<item type="delim" pos="41"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="42"><f>des</f><c>PRP_det</c><l>du</l></item>
<item type="delim" pos="43"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="44"><f>haillons</f><c>NOM</c><l>haillon</l></item>
<item type="delim" pos="45"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="delim" pos="46"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="47"><f>D</f><c>NAM</c><l>D</l></item>
<item type="delim" pos="48"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="delim" pos="49"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="50"><f>argent</f><c>NOM</c><l>argent</l></item>
<item type="delim" pos="51"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="delim" pos="52"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="delim" pos="53"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="54"><f>où</f><c>PRO_REL</c><l>où</l></item>
<item type="delim" pos="55"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="56"><f>le</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="57"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="58"><f>soleil</f><c>NOM</c><l>soleil</l></item>
<item type="delim" pos="59"><f> </f><c>DELIM</c><l>BLANK</l></item>
```

```
<item type="delim" pos="60"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="delim" pos="61"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="62"><f>de</f><c>PRP</c><l>de</l></item>
<item type="delim" pos="63"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="64"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="65"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="66"><f>montagne</f><c>NOM</c><l>montagne</l></item>
<item type="delim" pos="67"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="68"><f>fière</f><c>ADJ</c><l>fier</l></item>
<item type="delim" pos="69"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="70"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="delim" pos="71"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="72"><f>Luit</f><c>NAM</c><l>Luit</l></item>
<item type="delim" pos="73"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="delim" pos="74"><f>:</f><c>DELIM</c><l>:</l></item>
<item type="delim" pos="75"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="76"><f>c</f><c>VER_simp</c><l>c</l></item>
<item type="delim" pos="77"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="delim" pos="78"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="79"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="delim" pos="80"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="81"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="delim" pos="82"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="83"><f>petit</f><c>ADJ</c><l>petit</l></item>
<item type="delim" pos="84"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="85"><f>val</f><c>NOM</c><l>val</l></item>
<item type="delim" pos="86"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="87"><f>qui</f><c>PRO_REL</c><l>qui</l></item>
<item type="delim" pos="88"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="89"><f>mousse</f><c>NOM</c><l>mousse</l></item>
<item type="delim" pos="90"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="91"><f>de</f><c>PRP</c><l>de</l></item>
<item type="delim" pos="92"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="93"><f>rayons</f><c>NOM</c><l>rayon</l></item>
<item type="delim" pos="94"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="delim" pos="95"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="96"><f>Un</f><c>NAM</c><l>Un</l></item>
<item type="delim" pos="97"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="98"><f>soldat</f><c>NOM</c><l>soldat</l></item>
<item type="delim" pos="99"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="100"><f>jeune</f><c>ADJ</c><l>jeune</l></item>
<item type="delim" pos="101"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="102"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="103"><f>bouche</f><c>VER_pres</c><l>boucher</l></item>
<item type="delim" pos="104"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="105"><f>ouverte</f><c>ADJ</c><l>ouvert</l></item>
<item type="delim" pos="106"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="107"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="108"><f>tête</f><c>NOM</c><l>tête</l></item>
<item type="delim" pos="109"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="110"><f>nue</f><c>VER_pres</c><l>nuer</l></item>
<item type="delim" pos="111"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="112"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="113"><f>Et</f><c>NAM</c><l>Et</l></item>
<item type="delim" pos="114"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="115"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="116"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="117"><f>nuque</f><c>NOM</c><l>nuque</l></item>
<item type="delim" pos="118"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="119"><f>baignant</f><c>VER_ppre</c><l>baigner</l></item>
<item type="delim" pos="120"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="121"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="delim" pos="122"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="123"><f>le</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="124"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="125"><f>frais</f><c>ADJ</c><l>frais</l></item>
<item type="delim" pos="126"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="127"><f>cresson</f><c>NOM</c><l>cresson</l></item>
<item type="delim" pos="128"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="129"><f>bleu</f><c>ADJ</c><l>bleu</l></item>
<item type="delim" pos="130"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="131"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="132"><f>Dort</f><c>NAM</c><l>Dort</l></item>
<item type="delim" pos="133"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="delim" pos="134"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="delim" pos="135"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="136"><f>il</f><c>PRO_PER</c><l>il</l></item>
```

```
<item type="delim" pos="137"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="138"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="delim" pos="139"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="140"><f>étendu</f><c>VER_pper</c><l>étendre</l></item>
<item type="delim" pos="141"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="142"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="delim" pos="143"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="144"><f>l</f><c>VER_simp</c><l>l</l></item>
<item type="delim" pos="145"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="forme" pos="146"><f>herbe</f><c>NOM</c><l>herbe</l></item>
<item type="delim" pos="147"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="148"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="149"><f>sous</f><c>PRP</c><l>sous</l></item>
<item type="delim" pos="150"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="151"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="152"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="153"><f>nue</f><c>VER_pres</c><l>nuer</l></item>
<item type="delim" pos="154"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="155"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="156"><f>Pâle</f><c>NAM</c><l>Pâle</l></item>
<item type="delim" pos="157"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="158"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="delim" pos="159"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="160"><f>son</f><c>DET_POS</c><l>son</l></item>
<item type="delim" pos="161"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="162"><f>lit</f><c>VER_pres</c><l>lire</l></item>
<item type="delim" pos="163"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="164"><f>vert</f><c>ADJ</c><l>vert</l></item>
<item type="delim" pos="165"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="166"><f>où</f><c>PRO_REL</c><l>où</l></item>
<item type="delim" pos="167"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="168"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="169"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="170"><f>lumière</f><c>NOM</c><l>lumière</l></item>
<item type="delim" pos="171"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="172"><f>pleut</f><c>VER_pres</c><l>pleuvoir</l></item>
<item type="delim" pos="173"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="delim" pos="174"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="175"><f>Les</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="176"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="177"><f>pieds</f><c>NOM</c><l>pied</l></item>
<item type="delim" pos="178"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="179"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="delim" pos="180"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="181"><f>les</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="182"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="183"><f>glaïeuls</f><c>NOM</c><l>glaïeul</l></item>
<item type="delim" pos="184"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="185"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="186"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="delim" pos="187"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="188"><f>dort</f><c>VER_pres</c><l>dormir</l></item>
<item type="delim" pos="189"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="delim" pos="190"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="191"><f>Souriant</f><c>NAM</c><l>Souriant</l></item>
<item type="delim" pos="192"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="193"><f>comme</f><c>KON</c><l>comme</l></item>
<item type="delim" pos="194"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="195"><f>Sourirait</f><c>NAM</c><l>Sourirait</l></item>
<item type="delim" pos="196"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="197"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="delim" pos="198"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="199"><f>enfant</f><c>NOM</c><l>enfant</l></item>
<item type="delim" pos="200"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="201"><f>malade</f><c>ADJ</c><l>malade</l></item>
<item type="delim" pos="202"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="203"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="204"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="delim" pos="205"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="206"><f>fait</f><c>VER_pres</c><l>faire</l></item>
<item type="delim" pos="207"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="208"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="delim" pos="209"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="210"><f>somme</f><c>NOM</c><l>somme</l></item>
<item type="delim" pos="211"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="delim" pos="212"><f>:</f><c>DELIM</c><l>:</l></item>
<item type="delim" pos="213"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
```

```
<item type="forme" pos="214"><f>Nature</f><c>NAM</c><l>Nature</l></item>
<item type="delim" pos="215"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="216"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="217"><f>berce</f><c>VER_pres</c><l>bercer</l></item>
<item type="delim" pos="218"><f>-</f><c>DELIM</c><l>-</l></item>
<item type="forme" pos="219"><f>le</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="220"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="221"><f>chaudement</f><c>ADV</c><l>chaudement</l></item>
<item type="delim" pos="222"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="delim" pos="223"><f>:</f><c>DELIM</c><l>:</l></item>
<item type="delim" pos="224"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="225"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="delim" pos="226"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="227"><f>a</f><c>VER_pres</c><l>avoir</l></item>
<item type="delim" pos="228"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="229"><f>froid</f><c>ADJ</c><l>froid</l></item>
<item type="delim" pos="230"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="delim" pos="231"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="232"><f>Les</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="233"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="234"><f>parfums</f><c>NOM</c><l>parfum</l></item>
<item type="delim" pos="235"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="236"><f>ne</f><c>ADV</c><l>ne</l></item>
<item type="delim" pos="237"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="238"><f>font</f><c>VER_pres</c><l>faire</l></item>
<item type="delim" pos="239"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="240"><f>pas</f><c>ADV</c><l>pas</l></item>
<item type="delim" pos="241"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="242"><f>frissonner</f><c>VER_infi</c><l>frissonner</l></item>
<item type="delim" pos="243"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="244"><f>sa</f><c>DET_POS</c><l>son</l></item>
<item type="delim" pos="245"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="246"><f>narine</f><c>NOM</c><l>narine</l></item>
<item type="delim" pos="247"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="delim" pos="248"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="delim" pos="249"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="250"><f>Il</f><c>NAM</c><l>Il</l></item>
<item type="delim" pos="251"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="252"><f>dort</f><c>VER_pres</c><l>dormir</l></item>
<item type="delim" pos="253"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="254"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="delim" pos="255"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="256"><f>le</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="257"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="258"><f>soleil</f><c>NOM</c><l>soleil</l></item>
<item type="delim" pos="259"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="260"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="261"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="262"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="263"><f>main</f><c>NOM</c><l>main</l></item>
<item type="delim" pos="264"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="265"><f>sur</f><c>PRP</c><l>sur</l></item>
<item type="delim" pos="266"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="267"><f>sa</f><c>DET_POS</c><l>son</l></item>
<item type="delim" pos="268"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="269"><f>poitrine</f><c>NOM</c><l>poitrine</l></item>
<item type="delim" pos="270"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="271"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="272"><f>Tranquille</f><c>NAM</c><l>Tranquille</l></item>
<item type="delim" pos="273"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="delim" pos="274"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="275"><f>Il</f><c>NAM</c><l>Il</l></item>
<item type="delim" pos="276"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="277"><f>a</f><c>VER_pres</c><l>avoir</l></item>
<item type="delim" pos="278"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="279"><f>deux</f><c>NUM</c><l>deux</l></item>
<item type="delim" pos="280"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="281"><f>trous</f><c>NOM</c><l>trou</l></item>
<item type="delim" pos="282"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="283"><f>rouges</f><c>ADJ</c><l>rouge</l></item>
<item type="delim" pos="284"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="285"><f>au</f><c>PRP_det</c><l>au</l></item>
<item type="delim" pos="286"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="287"><f>côté</f><c>NOM</c><l>côté</l></item>
<item type="delim" pos="288"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="289"><f>droit</f><c>NOM</c><l>droit</l></item>
<item type="delim" pos="290"><f>.</f><c>DELIM</c><l>.</l></item>
```

```
<item type="delim" pos="291"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="forme" pos="292"><f>Arthur</f><c>NAM</c><l>Arthur</l></item>
<item type="delim" pos="293"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="294"><f>Rimbaud</f><c>NAM</c><l>Rimbaud</l></item>
<item type="delim" pos="295"><f>RETURN</f><c>DELIM</c><l>RETURN</l></item>
<item type="delim" pos="296"><f> </f><c>DELIM</c><l>BLANK</l></item>
</items>
</Trame>
<Cadre>
<acces>
<partition nom="ligne">
<p n="VERS1" d="10" f="33" nd="4" nf="5"/>
<p n="VERS2" d="33" f="46" nd="6" nf="7"/>
<p n="VERS3" d="46" f="71" nd="8" nf="9"/>
<p n="VERS4" d="71" f="96" nd="10" nf="11"/>
<p n="VERS5" d="96" f="113" nd="12" nf="13"/>
<p n="VERS6" d="113" f="132" nd="14" nf="15"/>
<p n="VERS7" d="132" f="156" nd="16" nf="17"/>
<p n="VERS8" d="156" f="175" nd="18" nf="19"/>
<p n="VERS9" d="175" f="195" nd="20" nf="21"/>
<p n="VERS10" d="195" f="214" nd="22" nf="23"/>
<p n="VERS11" d="214" f="232" nd="24" nf="25"/>
<p n="VERS12" d="232" f="250" nd="26" nf="27"/>
<p n="VERS13" d="250" f="272" nd="28" nf="29"/>
<p n="VERS14" d="272" f="292" nd="30" nf="31"/>
</partition>
<partition nom="structure">
<p n="TITRE" d="1" f="10" nd="1" nf="2"/>
<p n="TEXTEPOEME" d="10" f="292" nd="3" nf="32"/>
<p n="AUTEUR" d="292" f="296" nd="33" nf="34"/>
</partition>
</acces>
</Cadre>
</baselexicometrique>
```

Le fichier d'export contient tout d'abord la ***Trame*** (liste des positions des items et pour chacune d'elles les annotations disponibles : forme, lemme[22], catégorie[23], annotations importées) puis le ***Cadre*** (les intervalles de positions regroupant certaines unités de la ***Trame*** en parties structurées).

[22] Si treetagger a préalablement été activé
[23] Si treetagger a préalablement été activé

28 Exportation du *Cadre* et de la *Trame*

Il est possible d'exporter le ***Cadre*** et la ***Trame*** en tenant compte des éventuelles corrections apportées sur les mots, les lemmes ou les catégories ; de même les systèmes d'annotation importés (*cf infra*) seront intégrés dans les fichiers d'export.

En activant le bouton , le ***Trameur*** enregistre dans le dossier `export` 3 fichiers : la ***Trame***, le ***Cadre*** et la ***Trame*** annotée.

Nous allons illustrer sur un exemple les formats des fichiers d'export.

28.1 Fichier de travail

```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
<LIGNE="VERS5">Un soldat jeune, bouche ouverte, tête nue,
<LIGNE="VERS6">Et la nuque baignant dans le frais cresson bleu,
<LIGNE="VERS7">Dort ; il est étendu dans l'herbe, sous la nue,
<LIGNE="VERS8">Pâle dans son lit vert où la lumière pleut.
<LIGNE="VERS9">Les pieds dans les glaïeuls, il dort. Souriant comme
<LIGNE="VERS10">Sourirait un enfant malade, il fait un somme :
<LIGNE="VERS11">Nature, berce-le chaudement : il a froid.
<LIGNE="VERS12">Les parfums ne font pas frissonner sa narine ;
<LIGNE="VERS13">Il dort dans le soleil, la main sur sa poitrine,
<LIGNE="VERS14">Tranquille. Il a deux trous rouges au côté droit.
<STRUCTURE="AUTEUR">Arthur Rimbaud
```

28.2 Fichiers d'export

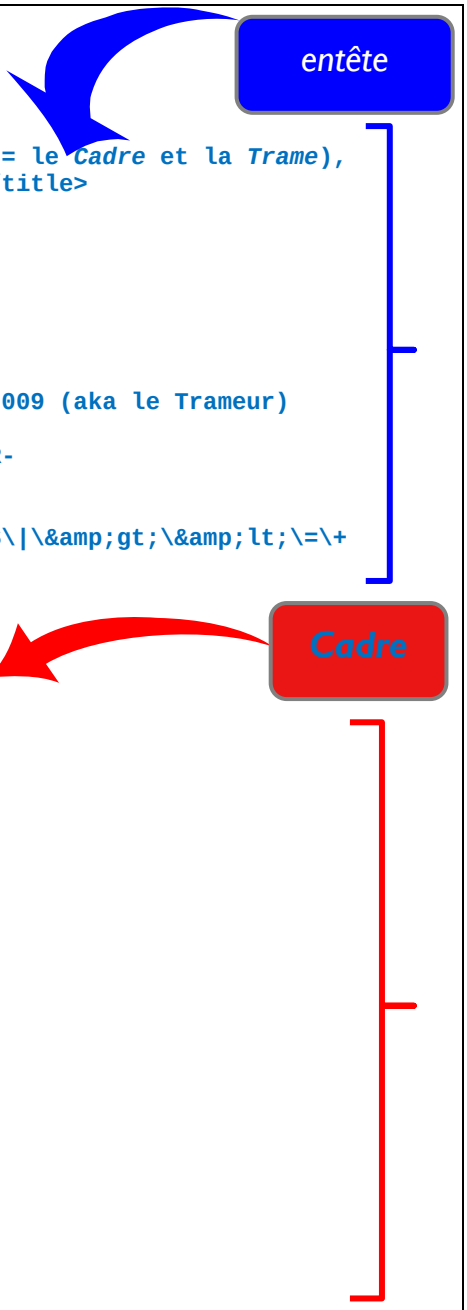
1. La *Trame* reconstituée :

```
  Le dormeur du val
 C' est un trou de verdure où chante une rivière,
 Accrochant follement aux herbes des haillons
 D' argent ; où le soleil ; de la montagne fière,
 Luit : c' est un petit val qui mousse de rayons.
Un soldat jeune, bouche ouverte, tête nue,
Et la nuque baignant dans le frais cresson bleu,
Dort ; il est étendu dans l'herbe, sous la nue,
Pâle dans son lit vert où la lumière pleut.
Les pieds dans les glaïeuls, il dort. Souriant comme
Sourirait un enfant malade, il fait un somme :
Nature, berce-le chaudement : il a froid.
Les parfums ne font pas frissonner sa narine ;
Il dort dans le soleil, la main sur sa poitrine,
Tranquille. Il a deux trous rouges au côté droit.
Arthur Rimbaud
```

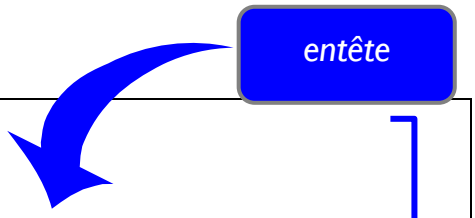
2. Le *Cadre* :

```
<?xml version="1.0" encoding="iso-8859-1"?>
<Cadre>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame),
importable et echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
<p>Samedi 26 Mai 2007
 10:50:52
  Ce document n'est pas encore publie.</p>
</publicationStmt>
<sourceDesc>
Fichier d'export genere par LeMetierLexicometrique 5.009 (aka le Trameur)
(http://www.tal.univ-paris3.fr/trameur/).
Fichier initilement traite : C:/SFleury/DEVEL/WIP-ANR-
textometrrie/mkGraphTrameCadre/textes/dormeurL3.txt
Encodage : iso-8859-1
Delimiteur : \.\ \,\:\;\!\?\/\_\-\"\'\(\)\[\]\{\}\§\$\|\&amp;gt;\&amp;lt;\=\+

</sourceDesc>
</fileDesc>
</teiheader>
<acces>
<partition nom="ligne">
<p n="VERS1" d="10" f="33" nd="4" nf="5"/>
<p n="VERS2" d="33" f="46" nd="6" nf="7"/>
<p n="VERS3" d="46" f="71" nd="8" nf="9"/>
<p n="VERS4" d="71" f="96" nd="10" nf="11"/>
<p n="VERS5" d="96" f="113" nd="12" nf="13"/>
<p n="VERS6" d="113" f="132" nd="14" nf="15"/>
<p n="VERS7" d="132" f="156" nd="16" nf="17"/>
<p n="VERS8" d="156" f="175" nd="18" nf="19"/>
<p n="VERS9" d="175" f="195" nd="20" nf="21"/>
<p n="VERS10" d="195" f="214" nd="22" nf="23"/>
<p n="VERS11" d="214" f="232" nd="24" nf="25"/>
<p n="VERS12" d="232" f="250" nd="26" nf="27"/>
<p n="VERS13" d="250" f="272" nd="28" nf="29"/>
<p n="VERS14" d="272" f="292" nd="30" nf="31"/>
</partition>
<partition nom="structure">
<p n="TITRE" d="1" f="10" nd="1" nf="2"/>
<p n="TEXTEPOEME" d="10" f="292" nd="3" nf="32"/>
<p n="AUTEUR" d="292" f="296" nd="33" nf="34"/>
</partition>
</acces>
</Cadre>
```



3. La *Trame* annotée :



```
<?xml version="1.0" encoding="iso-8859-1"?>
<trametreetagger>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame), importable et
echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
<p>Samedi 26 Mai 2007
 10:41:33
  Ce document n'est pas encore publie.</p>
</publicationStmt>
<sourceDesc>
Fichier d'export genere par LeMetierLexicometrique 5.009 (aka le Trameur)
(http://www.tal.univ-paris3.fr/trameur/).
Fichier initilement traite : C:/SFleury/DEVEL/WIP-ANR-
textometrrie/mkGraphTrameCadre/textes/dormeurL3.txt
Encodage : iso-8859-1
Delimiteur : \.\ \,\:\;\!\?\/\_\-\"\'\(\)\[\]\{\}\§\$\|\&amp;gt;\&amp;lt;\=\+

</sourceDesc>
</fileDesc>
</teiheader>
<items>
<item type="forme" pos="1"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="2"><f>Le</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="3"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="4"><f>dormeur</f><c>NOM</c><l>dormeur</l></item>
<item type="forme" pos="5"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="6"><f>du</f><c>PRP_det</c><l>du</l></item>
<item type="forme" pos="7"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="8"><f>val</f><c>NOM</c><l>val</l></item>
<item type="forme" pos="9"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="10"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="11"><f>C</f><c>NUM</c><l>C</l></item>
<item type="forme" pos="12"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="forme" pos="13"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="14"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="forme" pos="15"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="16"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="forme" pos="17"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="18"><f>trou</f><c>NOM</c><l>trou</l></item>
<item type="forme" pos="19"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="20"><f>de</f><c>PRP</c><l>de</l></item>
<item type="forme" pos="21"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="22"><f>verdure</f><c>NOM</c><l>verdure</l></item>
<item type="forme" pos="23"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="24"><f>où</f><c>PRO_REL</c><l>où</l></item>
<item type="forme" pos="25"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="26"><f>chante</f><c>VER_pres</c><l>chanter</l></item>
<item type="forme" pos="27"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="28"><f>une</f><c>DET_ART</c><l>un</l></item>
<item type="forme" pos="29"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="30"><f>rivière</f><c>NOM</c><l>rivière</l></item>
<item type="forme" pos="31"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="32"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="33"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="34"><f>Accrochant</f><c>NAM</c><l>Accrochant</l></item>
<item type="forme" pos="35"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="36"><f>follement</f><c>ADV</c><l>follement</l></item>
<item type="forme" pos="37"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="38"><f>aux</f><c>PRP_det</c><l>au</l></item>
<item type="forme" pos="39"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="40"><f>herbes</f><c>VER_pres</c><l>herber</l></item>
<item type="forme" pos="41"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="42"><f>des</f><c>PRP_det</c><l>du</l></item>
<item type="forme" pos="43"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="44"><f>haillons</f><c>NOM</c><l>haillon</l></item>
<item type="forme" pos="45"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="46"><f> </f><c>BLANK</c><l>BLANK</l></item>
```

```
<item type="forme" pos="47"><f>D</f><c>NAM</c><l>D</l></item>
<item type="forme" pos="48"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="forme" pos="49"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="50"><f>argent</f><c>NOM</c><l>argent</l></item>
<item type="forme" pos="51"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="52"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="forme" pos="53"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="54"><f>où</f><c>PRO_REL</c><l>où</l></item>
<item type="forme" pos="55"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="56"><f>le</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="57"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="58"><f>soleil</f><c>NOM</c><l>soleil</l></item>
<item type="forme" pos="59"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="60"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="forme" pos="61"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="62"><f>de</f><c>PRP</c><l>de</l></item>
<item type="forme" pos="63"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="64"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="65"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="66"><f>montagne</f><c>NOM</c><l>montagne</l></item>
<item type="forme" pos="67"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="68"><f>fière</f><c>ADJ</c><l>fier</l></item>
<item type="forme" pos="69"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="70"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="71"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="72"><f>Luit</f><c>NAM</c><l>Luit</l></item>
<item type="forme" pos="73"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="74"><f>:</f><c>DELIM</c><l>:</l></item>
<item type="forme" pos="75"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="76"><f>c</f><c>VER_simp</c><l>c</l></item>
<item type="forme" pos="77"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="forme" pos="78"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="79"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="forme" pos="80"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="81"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="forme" pos="82"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="83"><f>petit</f><c>ADJ</c><l>petit</l></item>
<item type="forme" pos="84"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="85"><f>val</f><c>NOM</c><l>val</l></item>
<item type="forme" pos="86"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="87"><f>qui</f><c>PRO_REL</c><l>qui</l></item>
<item type="forme" pos="88"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="89"><f>mousse</f><c>NOM</c><l>mousse</l></item>
<item type="forme" pos="90"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="91"><f>de</f><c>PRP</c><l>de</l></item>
<item type="forme" pos="92"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="93"><f>rayons</f><c>NOM</c><l>rayon</l></item>
<item type="forme" pos="94"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="forme" pos="95"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="96"><f>Un</f><c>NAM</c><l>Un</l></item>
<item type="forme" pos="97"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="98"><f>soldat</f><c>NOM</c><l>soldat</l></item>
<item type="forme" pos="99"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="100"><f>jeune</f><c>ADJ</c><l>jeune</l></item>
<item type="forme" pos="101"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="102"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="103"><f>bouche</f><c>VER_pres</c><l>boucher</l></item>
<item type="forme" pos="104"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="105"><f>ouverte</f><c>ADJ</c><l>ouvert</l></item>
<item type="forme" pos="106"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="107"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="108"><f>tête</f><c>NOM</c><l>tête</l></item>
<item type="forme" pos="109"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="110"><f>nue</f><c>VER_pres</c><l>nuer</l></item>
<item type="forme" pos="111"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="112"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="113"><f>Et</f><c>NAM</c><l>Et</l></item>
<item type="forme" pos="114"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="115"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="116"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="117"><f>nuque</f><c>NOM</c><l>nuque</l></item>
<item type="forme" pos="118"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="119"><f>baignant</f><c>VER_ppre</c><l>baigner</l></item>
<item type="forme" pos="120"><f> </f><c>BLANK</c><l>BLANK</l></item>
```

```
<item type="forme" pos="121"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="forme" pos="122"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="123"><f>le</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="124"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="125"><f>frais</f><c>ADJ</c><l>frais</l></item>
<item type="forme" pos="126"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="127"><f>cresson</f><c>NOM</c><l>cresson</l></item>
<item type="forme" pos="128"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="129"><f>bleu</f><c>ADJ</c><l>bleu</l></item>
<item type="forme" pos="130"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="131"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="132"><f>Dort</f><c>NAM</c><l>Dort</l></item>
<item type="forme" pos="133"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="134"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="forme" pos="135"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="136"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="forme" pos="137"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="138"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="forme" pos="139"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="140"><f>étendu</f><c>VER_pper</c><l>étendre</l></item>
<item type="forme" pos="141"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="142"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="forme" pos="143"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="144"><f>l</f><c>VER_simp</c><l>l</l></item>
<item type="forme" pos="145"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="forme" pos="146"><f>herbe</f><c>NOM</c><l>herbe</l></item>
<item type="forme" pos="147"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="148"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="149"><f>sous</f><c>PRP</c><l>sous</l></item>
<item type="forme" pos="150"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="151"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="152"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="153"><f>nue</f><c>VER_pres</c><l>nuer</l></item>
<item type="forme" pos="154"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="155"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="156"><f>Pâle</f><c>NAM</c><l>Pâle</l></item>
<item type="forme" pos="157"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="158"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="forme" pos="159"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="160"><f>son</f><c>DET_POS</c><l>son</l></item>
<item type="forme" pos="161"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="162"><f>lit</f><c>VER_pres</c><l>lire</l></item>
<item type="forme" pos="163"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="164"><f>vert</f><c>ADJ</c><l>vert</l></item>
<item type="forme" pos="165"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="166"><f>où</f><c>PRO_REL</c><l>où</l></item>
<item type="forme" pos="167"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="168"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="169"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="170"><f>lumière</f><c>NOM</c><l>lumière</l></item>
<item type="forme" pos="171"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="172"><f>pleut</f><c>VER_pres</c><l>pleuvoir</l></item>
<item type="forme" pos="173"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="forme" pos="174"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="175"><f>Les</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="176"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="177"><f>pieds</f><c>NOM</c><l>pied</l></item>
<item type="forme" pos="178"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="179"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="forme" pos="180"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="181"><f>les</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="182"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="183"><f>glaïeuls</f><c>NOM</c><l>glaïeul</l></item>
<item type="forme" pos="184"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="185"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="186"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="forme" pos="187"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="188"><f>dort</f><c>VER_pres</c><l>dormir</l></item>
<item type="forme" pos="189"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="forme" pos="190"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="191"><f>Souriant</f><c>NAM</c><l>Souriant</l></item>
<item type="forme" pos="192"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="193"><f>comme</f><c>KON</c><l>comme</l></item>
<item type="forme" pos="194"><f>
```

```
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="195"><f>Sourirait</f><c>NAM</c><l>Sourirait</l></item>
<item type="forme" pos="196"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="197"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="forme" pos="198"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="199"><f>enfant</f><c>NOM</c><l>enfant</l></item>
<item type="forme" pos="200"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="201"><f>malade</f><c>ADJ</c><l>malade</l></item>
<item type="forme" pos="202"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="203"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="204"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="forme" pos="205"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="206"><f>fait</f><c>VER_pres</c><l>faire</l></item>
<item type="forme" pos="207"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="208"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="forme" pos="209"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="210"><f>somme</f><c>NOM</c><l>somme</l></item>
<item type="forme" pos="211"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="212"><f>:</f><c>DELIM</c><l>:</l></item>
<item type="forme" pos="213"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="214"><f>Nature</f><c>NAM</c><l>Nature</l></item>
<item type="forme" pos="215"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="216"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="217"><f>berce</f><c>VER_pres</c><l>bercer</l></item>
<item type="forme" pos="218"><f>-</f><c>DELIM</c><l>-</l></item>
<item type="forme" pos="219"><f>le</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="220"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="221"><f>chaudement</f><c>ADV</c><l>chaudement</l></item>
<item type="forme" pos="222"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="223"><f>:</f><c>DELIM</c><l>:</l></item>
<item type="forme" pos="224"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="225"><f>il</f><c>PRO_PER</c><l>il</l></item>
<item type="forme" pos="226"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="227"><f>a</f><c>VER_pres</c><l>avoir</l></item>
<item type="forme" pos="228"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="229"><f>froid</f><c>ADJ</c><l>froid</l></item>
<item type="forme" pos="230"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="forme" pos="231"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="232"><f>Les</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="233"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="234"><f>parfums</f><c>NOM</c><l>parfum</l></item>
<item type="forme" pos="235"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="236"><f>ne</f><c>ADV</c><l>ne</l></item>
<item type="forme" pos="237"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="238"><f>font</f><c>VER_pres</c><l>faire</l></item>
<item type="forme" pos="239"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="240"><f>pas</f><c>ADV</c><l>pas</l></item>
<item type="forme" pos="241"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="242"><f>frissonner</f><c>VER_infi</c><l>frissonner</l></item>
<item type="forme" pos="243"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="244"><f>sa</f><c>DET_POS</c><l>son</l></item>
<item type="forme" pos="245"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="246"><f>narine</f><c>NOM</c><l>narine</l></item>
<item type="forme" pos="247"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="248"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="forme" pos="249"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="250"><f>Il</f><c>NAM</c><l>Il</l></item>
<item type="forme" pos="251"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="252"><f>dort</f><c>VER_pres</c><l>dormir</l></item>
<item type="forme" pos="253"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="254"><f>dans</f><c>PRP</c><l>dans</l></item>
<item type="forme" pos="255"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="256"><f>le</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="257"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="258"><f>soleil</f><c>NOM</c><l>soleil</l></item>
<item type="forme" pos="259"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="260"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="261"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="forme" pos="262"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="263"><f>main</f><c>NOM</c><l>main</l></item>
<item type="forme" pos="264"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="265"><f>sur</f><c>PRP</c><l>sur</l></item>
<item type="forme" pos="266"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="267"><f>sa</f><c>DET_POS</c><l>son</l></item>
```

```
<item type="forme" pos="268"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="269"><f>poitrine</f><c>NOM</c><l>poitrine</l></item>
<item type="forme" pos="270"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="forme" pos="271"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="272"><f>Tranquille</f><c>NAM</c><l>Tranquille</l></item>
<item type="forme" pos="273"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="forme" pos="274"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="275"><f>Il</f><c>NAM</c><l>Il</l></item>
<item type="forme" pos="276"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="277"><f>a</f><c>VER_pres</c><l>avoir</l></item>
<item type="forme" pos="278"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="279"><f>deux</f><c>NUM</c><l>deux</l></item>
<item type="forme" pos="280"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="281"><f>trous</f><c>NOM</c><l>trou</l></item>
<item type="forme" pos="282"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="283"><f>rouges</f><c>ADJ</c><l>rouge</l></item>
<item type="forme" pos="284"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="285"><f>au</f><c>PRP_det</c><l>au</l></item>
<item type="forme" pos="286"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="287"><f>côté</f><c>NOM</c><l>côté</l></item>
<item type="forme" pos="288"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="289"><f>droit</f><c>NOM</c><l>droit</l></item>
<item type="forme" pos="290"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="forme" pos="291"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="292"><f>Arthur</f><c>NAM</c><l>Arthur</l></item>
<item type="forme" pos="293"><f> </f><c>BLANK</c><l>BLANK</l></item>
<item type="forme" pos="294"><f>Rimbaud</f><c>NAM</c><l>Rimbaud</l></item>
<item type="forme" pos="295"><f>
</f><c>RETURN</c><l>RETURN</l></item>
<item type="forme" pos="296"><f> </f><c>BLANK</c><l>BLANK</l></item>
</items>
</trametreetagger>
```

29 Le rapport : enregistrement des résultats produits

Les résultats qui intéressent l'utilisateur pour une exploitation ultérieure peuvent être rassemblés dans un dossier construit dans le répertoire `rapport`. Ce dossier aisément manipulable à l'aide d'un navigateur web (*Firefox*, *Internet Explorer*, *Safari*, etc.) contient un fichier d'index qui permet la navigation parmi les résultats sélectionnés. Le rapport peut être consulté dès que l'utilisateur l'a enregistré.

29.1 Ajouter un élément au rapport

Pour ajouter un document au rapport, il suffit de se positionner sur un onglet (ou une fenêtre) contenant un résultat produit par le programme puis de cliquer sur l'icône *Ajouter au rapport* présente dans la fenêtre principale ou dans la fenêtre de résultats (c'est le cas pour les spécificités).

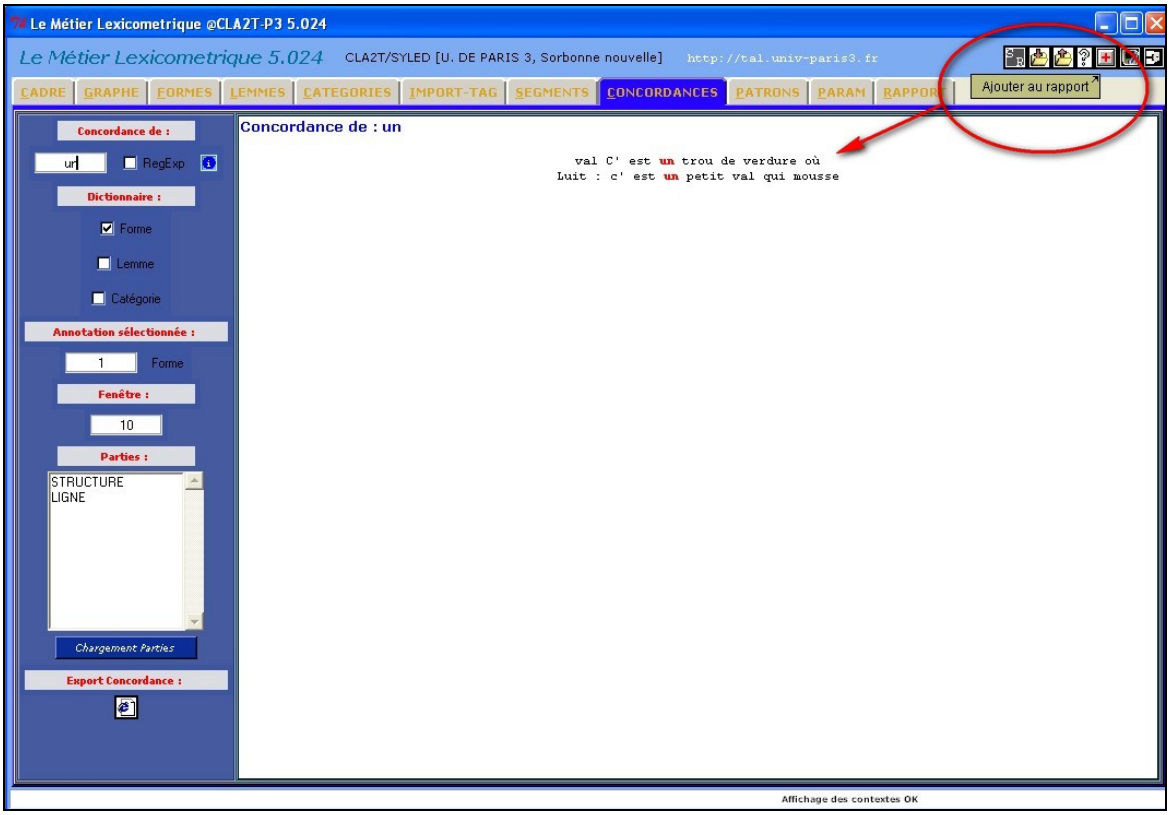


Figure 234 : Ajouter un élément au rapport

Une trace des éléments ajoutés au rapport est visible de manière synthétique dans l'onglet **RAPPORT**.

Le rapport est constitué par un cartouche et la liste des éléments ajoutés par l'utilisateur. Chacun des éléments du rapport dispose d'une zone de texte « libre » dans laquelle l'utilisateur peut y noter ses observations. Pour éditer ou modifier cette note, il suffit d'activer l'icône de l'éditeur présent près de chaque élément du rapport dans l'onglet **RAPPORT**.

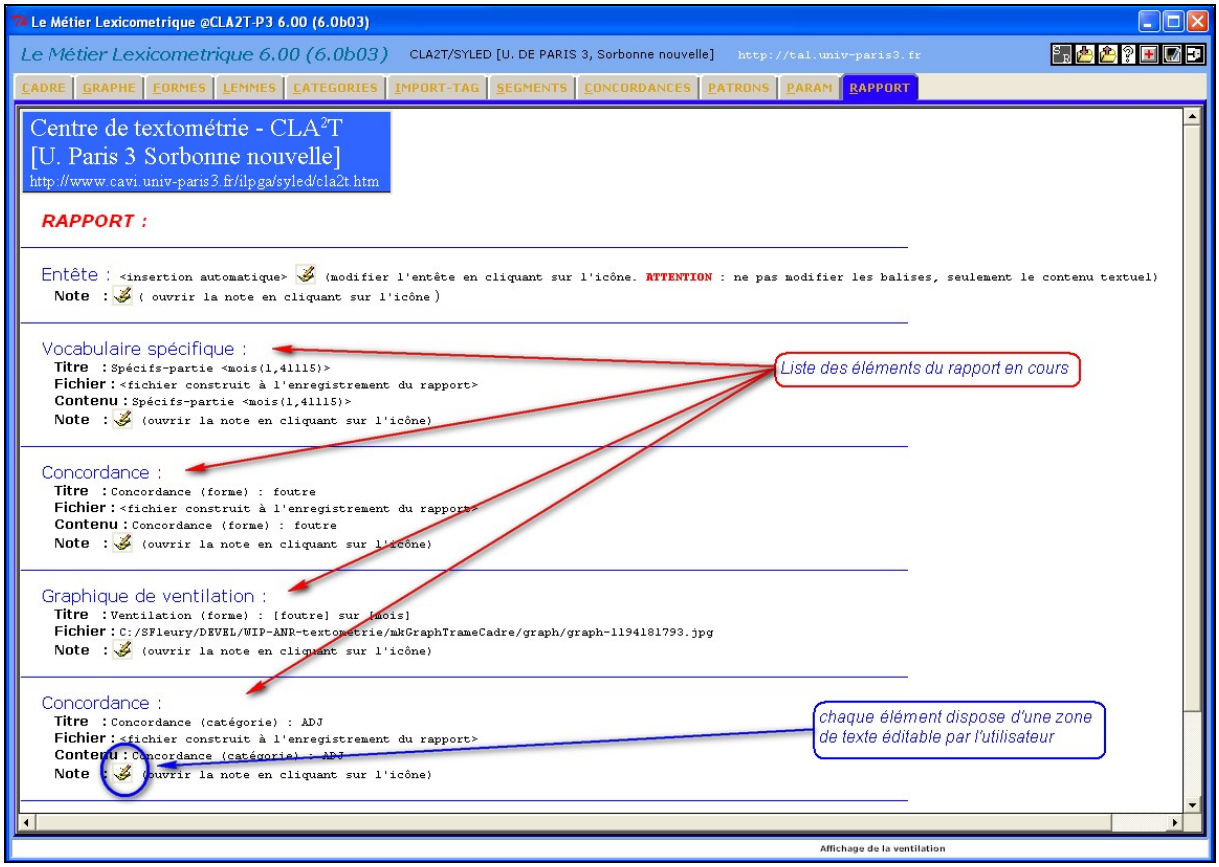


Figure 235 : Rapport en cours dans l'onglet RAPPORT

29.2 Ajouter un élément externe au rapport

Cette fonctionnalité permet à l'utilisateur d'insérer dans le rapport des images externes (non produites directement par le ***Trameur***). Par exemple, si l'on souhaite insérer une image donnant à voir la carte des sections, on peut le faire en procédant de la manière suivante :

- créer une copie d'écran de la carte des sections (par exemple avec un outil comme *FastStone Capture* http://www.faststone.org/) et sauvegarder la copie d'écran au format `jpeg` ou `gif`
- insérer l'image dans le rapport en utilisant le menu idoine dans l'onglet **RAPPORT**

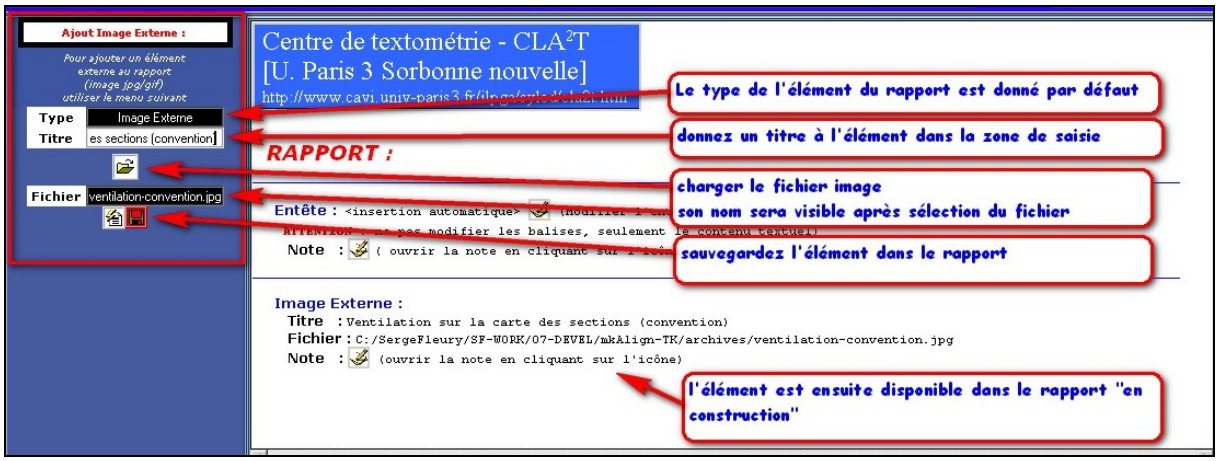

Ajout Image Externe :
Pour ajouter un élément externe au rapport (image jpg/gif) utiliser le menu suivant
Type
Image Externe
Titre
es sections (convention)
Fichier
ventilation-convention.jpg
Centre de textométrie - CLA²T
[U. Paris 3 Sorbonne nouvelle]
RAPPORT :
Entête : <insertion automatique>
Note :
Le type de l'élément du rapport est donné par défaut
donnez un titre à l'élément dans la zone de saisie
charger le fichier image
son nom sera visible après sélection du fichier
sauvegardez l'élément dans le rapport
Image Externe :
Titre : Ventilation sur la carte des sections (convention)
Fichier : C:/SergeFleury/SF-WORK/07-DEVEL/mkAlign-TK/archives/ventilation-convention.jpg
Note : (ouvrir la note en cliquant sur l'icône)
l'élément est ensuite disponible dans le rapport "en construction"

Figure 236 : Ajouter un élément externe au rapport

29.2 Enregistrer le rapport

On peut à tout moment visualiser le rapport en cliquant sur le bouton *Enregistrer le rapport* . Une fois activé, ce bouton déclenche la génération d'un « dossier rapport ». Ce dossier contient, dans la distribution initiale du ***Trameur***, des feuilles de styles XSL (pour chaque type de résultat à afficher) et une feuille de styles CSS (pour gérer l'affichage général des toutes les pages du rapport). Les fichiers construits pour un rapport sont au format XML et ils sont rassemblés dans un sous-dossier horodaté du dossier `rapport`.

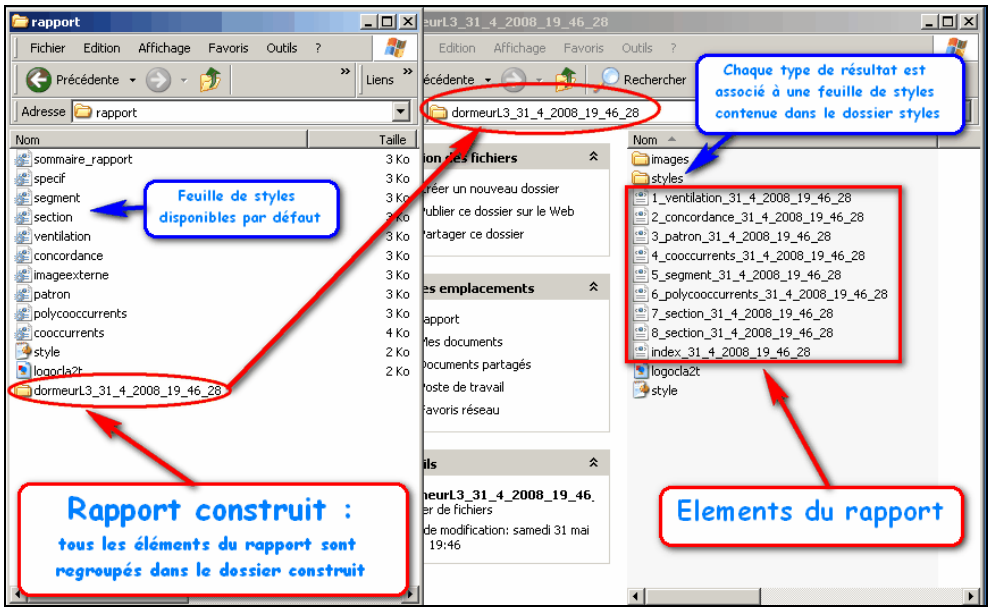
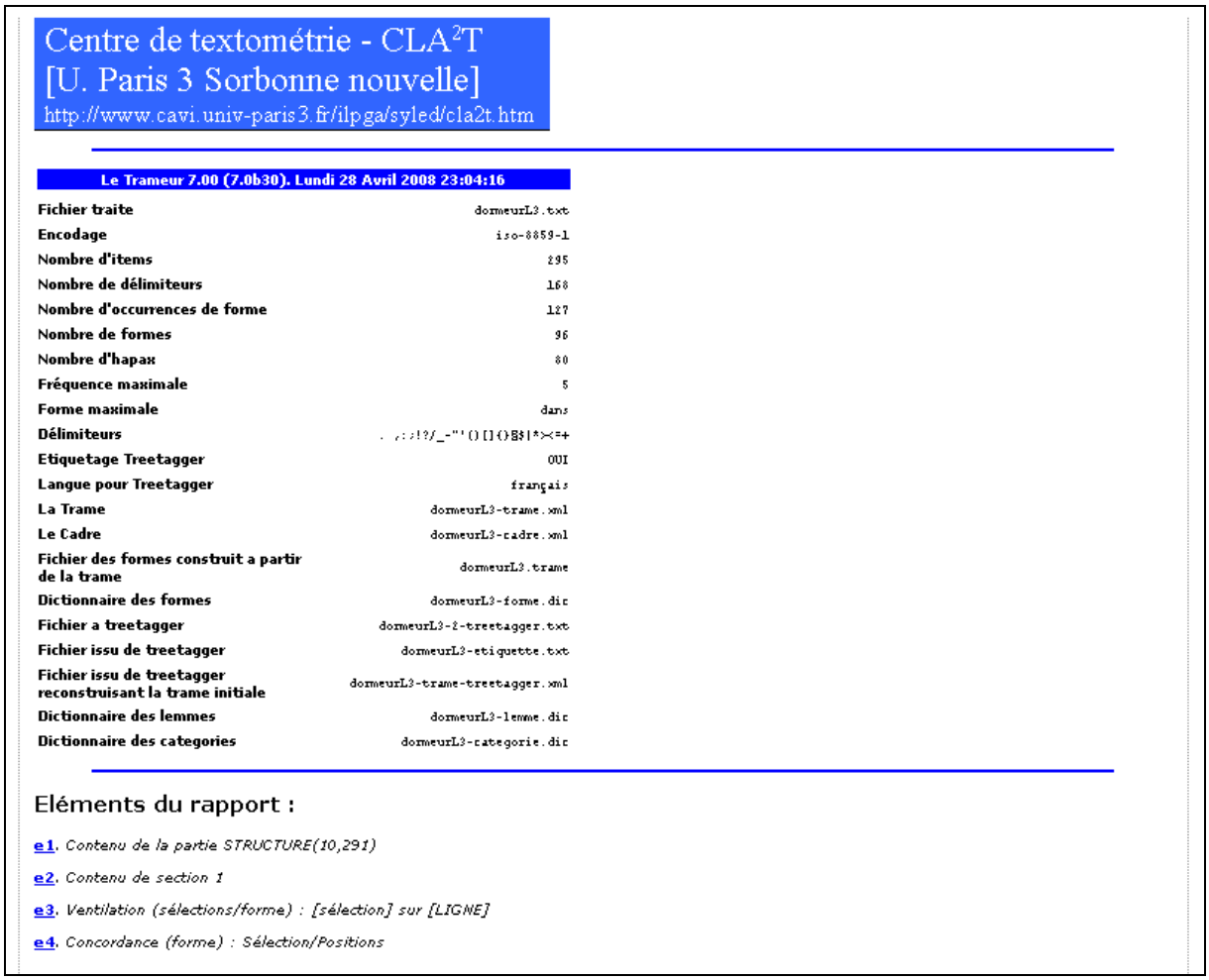


Figure 237 : Rapport enregistré

Après enregistrement, le rapport apparaît dans le navigateur web paramétré par défaut sur la machine utilisée (ici *Firefox*) :



Centre de textométrie - CLA²T
[U. Paris 3 Sorbonne nouvelle]
http://www.cavi.univ-paris3.fr/ilpga/syled/cla2t.htm

Le Trameur 7.00 (7.0b30). Lundi 28 Avril 2008 23:04:16

Fichier traite	dormeurL3.txt
Encodage	iso-8859-1
Nombre d'items	295
Nombre de délimiteurs	168
Nombre d'occurrences de forme	127
Nombre de formes	96
Nombre d'hapax	80
Fréquence maximale	5
Forme maximale	dans
Délimiteurs	. ,:;!?/_-"'()[]{}§$|*><=+
Etiquetage Treetagger	OUI
Langue pour Treetagger	français
La Trame	dormeurL3-trame.xml
Le Cadre	dormeurL3-cadre.xml
Fichier des formes construit a partir de la trame	dormeurL3.trame
Dictionnaire des formes	dormeurL3-forme.dic
Fichier a treetagger	dormeurL3-2-treetagger.txt
Fichier issu de treetagger	dormeurL3-etiquette.txt
Fichier issu de treetagger reconstruisant la trame initiale	dormeurL3-trame-treetagger.xml
Dictionnaire des lemmes	dormeurL3-lemme.dic
Dictionnaire des categories	dormeurL3-categorie.dic

Eléments du rapport :

e1. *Contenu de la partie STRUCTURE(10,291)*

e2. *Contenu de section 1*

e3. *Ventilation (sélections/forme) : [sélection] sur [LIGNE]*

e4. *Concordance (forme) : Sélection/Positions*

Figure 238 : Index du rapport enregistré

Chaque élément du rapport contient un menu de navigation (à gauche) pour accéder à tous les résultats construits (les différents éléments du rapport) et (à droite) de la description de l'élément visé.

Cette description d'un élément du rapport (sur la droite) est constituée :

- d'un *corps* (le contenu de l'*élément du rapport* pour les fichiers résultats ou une liste des *éléments du rapport* pour le fichier d'index)
- d'une *note* : zone de texte libre renseignée par l'utilisateur

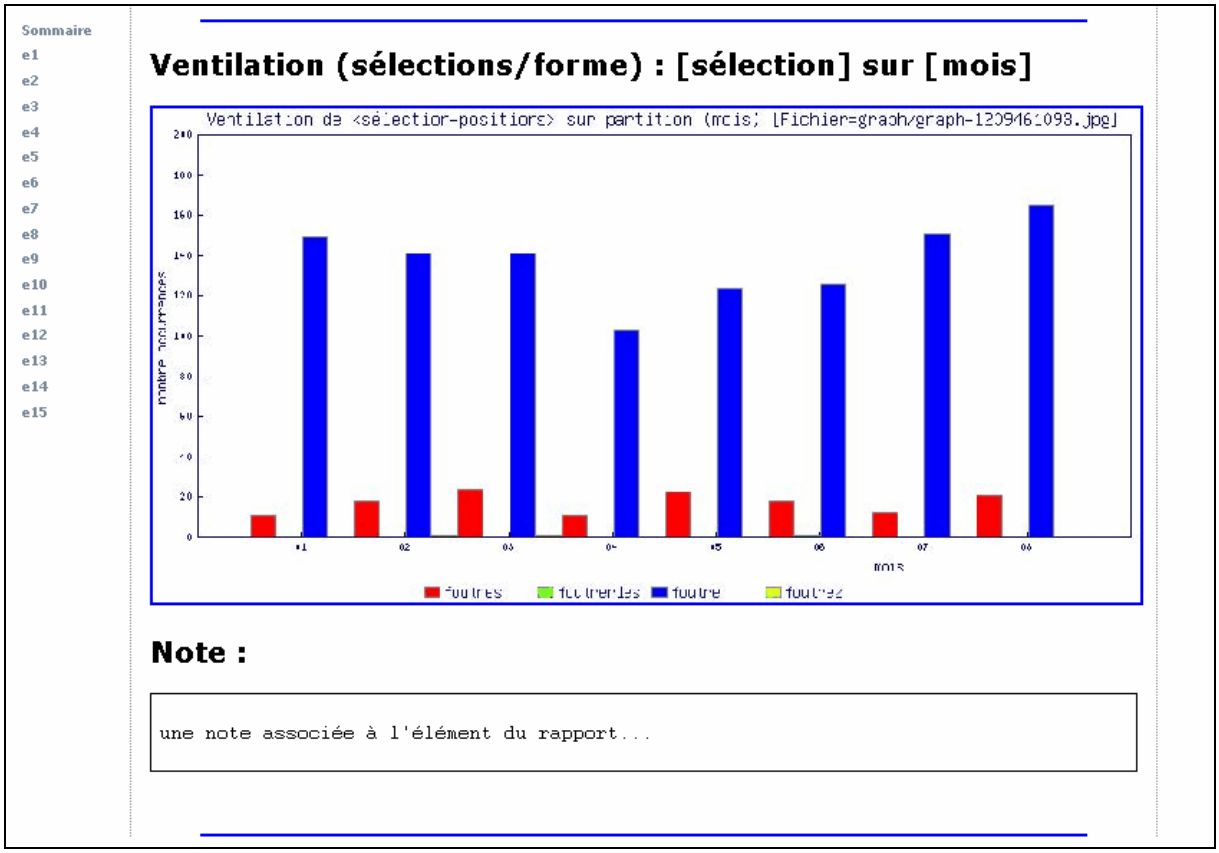


Figure 239 : Elément du rapport enregistré

On trouvera un exemple complet de rapport sur la page web du ***Trameur***.

30 Fonctionnalités complémentaires

30.1 Graphes d'annotation

30.1.1 Intégration dans le *Trameur* d'un module d'importation d'un graphe d'annotations sur la *Trame*

(*cf* « Perspective fonctionnelle de la phrase : la relation thème-rhème » F. Daoust 2009)

Illustration sur l'exemple de François :

> *Et un jour Jeanie partit à la recherche de son amoureux. Elle regardait les fleurs d' eau et leurs tiges penchées : et toutes les fleurs s' inclinaient vers elle. Et Jeanie disait en marchant ...*

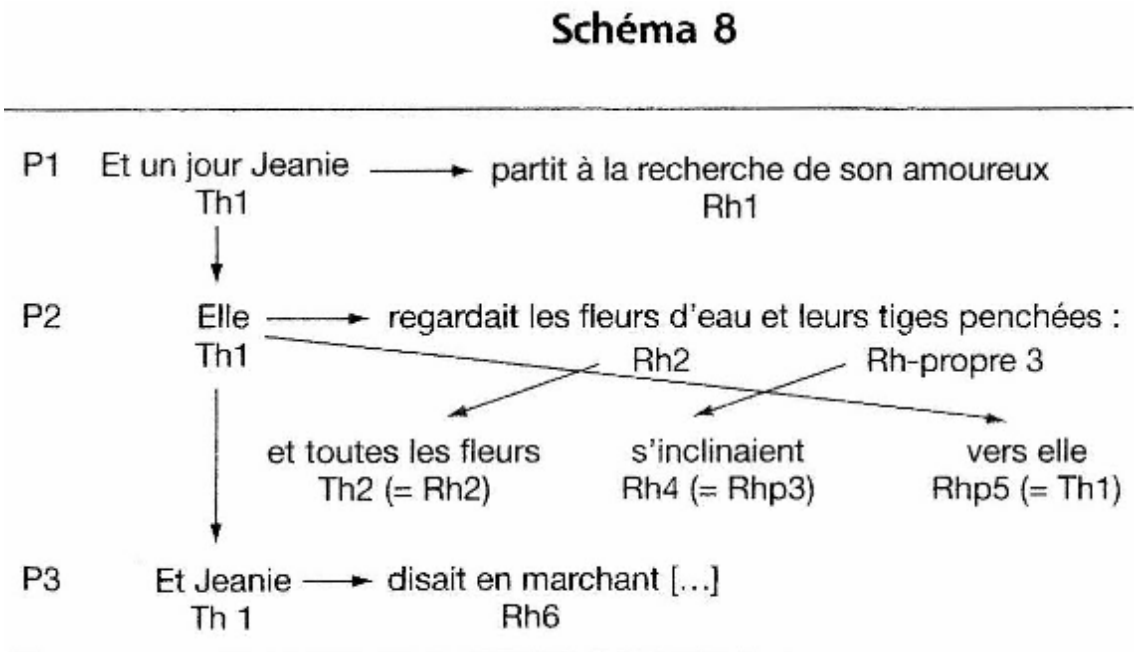
Les données de test utilisées ici (fichiers texte de travail et annotation+graphe) sont disponibles dans le dossier textes du répertoire de travail du ***Trameur*** après installation.

Etape 1

Ce texte est chargé dans le ***Trameur*** (création d'une base textométrique) : une ***Trame*** est créée, pas de ***Cadre*** explicite ici.

Le schéma suivant décrit le système d'annotation et son graphe (*cf* Daoust 2009):

Schéma 8



Le fichier suivant décrit un exemple de fichier à importer décrivant le système d'annotation (positions sur la ***Trame***) et le graphe associé :

```
<?xml version="1.0" encoding="iso-8859-1"?>
<TEI>
    <teiHeader>
        <fileDesc>
            <titleStmt>
                <title>Export Graphe Annotation sur la Trame d'un texte via Le Trameur
8.0b074</title>
            </titleStmt>
            <publicationStmt>
                <p>...</p>
            </publicationStmt>
            <sourceDesc>
                <bibl> ... </bibl>
            </sourceDesc>
        </fileDesc>
        <encodingDesc>
            <p>...</p>
        </encodingDesc>
    </teiHeader>
<text>
<body>
<graph type="directed" name="annotation-structurelle">
<node positions="1_3_5_7" labelnode="Th1"/>
<node positions="9_11_13_15_17_19_21" labelnode="Rh1"/>
<node positions="24" labelnode="Th1"/>
<node positions="26_28_30_32_35_37_39_41_43" labelnode="Rh2"/>
<node positions="37_39_41_43" labelnode="Rh-propre-3"/>
<node positions="47_49_51_53" labelnode="Th2"/>
<node positions="55_58" labelnode="Rh4"/>
<node positions="60_62" labelnode="Rhp5"/>
<node positions="65_67" labelnode="Th1"/>
<node positions="69_71_73" labelnode="Rh6"/>
<arc     from="1_3_5_7"     to="9_11_13_15_17_19_21"     labelfrom="Th1"     labelto="Rh1"
labelarc="dépendance à définir"/>
<arc   from="1_3_5_7"   to="24"   labelfrom="Th1"   labelto="Th1"   labelarc="dépendance   à
définir"/>
<arc from="24" to="65_67" labelfrom="Th1" labelto="Th1" labelarc="dépendance à définir"/>
<arc     from="24"     to="26_28_30_32_35_37_39_41_43"     labelfrom="Th1"     labelto="Rh2"
labelarc="dépendance à définir"/>
<arc   from="26_28_30_32_35_37_39_41_43"   to="37_39_41_43"   labelfrom="Rh2"   labelto="Rh-
propre-3" labelarc="dépendance à définir"/>
<arc   from="26_28_30_32_35_37_39_41_43"   to="47_49_51_53"   labelfrom="Rh2"   labelto="Th2"
labelarc="dépendance à définir"/>
<arc      from="37_39_41_43"      to="55_58"      labelfrom="Rh-propre-3"      labelto="Rh4"
labelarc="dépendance à définir"/>
<arc from="24" to="60_62" labelfrom="Th1" labelto="Rhp5" labelarc="dépendance à définir"/>
<arc  from="65_67"  to="69_71_73"  labelfrom="Th1"  labelto="Rh6"  labelarc="dépendance  à
définir"/>
</graph>
</body>
</text>
</TEI>
```

- Chaque élément `node` décrit les positions visées de la ***Trame*** (*via* l'attribut `positions`), et la valeur du label pour ce nœud de positions (*via* l'attribut `labelnode`). Quand le label du nœud n'est pas spécifié explicitement, une valeur par défaut lui est attribuée (les positions de ses constituants).
- Chaque élément `arc` décrit les liens entre les éléments `node` : l'attribut `from` porte les positions de la ***Trame*** du nœud source de l'arc, l'attribut `to` porte les positions de la ***Trame*** du nœud cible de l'arc, l'attribut `labelfrom` porte la valeur du label du nœud source, l'attribut `labelto` porte la valeur du label du nœud cible et l'attribut `labelarc` porte la valeur du label associé à l'arc.

Les positions de la ***Trame*** sont désignées en les listant une à une et en respectant ce format : `positions="26_27_35",` ici 3 positions discontinues de la ***Trame*** sont désignées.

Etape 2 : importation de l'annotation et du graphe associé dans Le Trameur

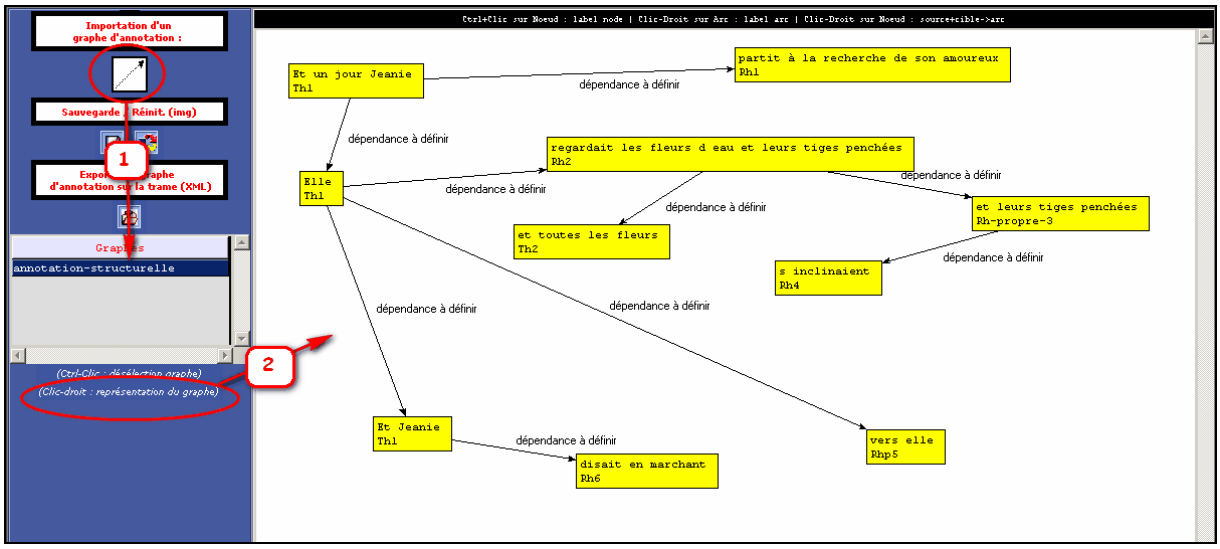


Figure 240 : Importation d'annotation+graphe

Une fois chargée, le graphe est disponible en mémoire dans le ***Trameur*** (1). On peut aussi visualiser le graphe de cette annotation (2).

30.1.1 Génération de graphes d'annotations

Nous allons montrer ci-dessous comment générer directement dans le ***Trameur*** un tel graphe en reconstruisant le graphe vu dans la partie précédente.

Etape 1

Le texte précédent est chargé dans le ***Trameur*** (création d'une base textométrique) : une ***Trame*** est créée, pas de ***Cadre*** explicite ici.

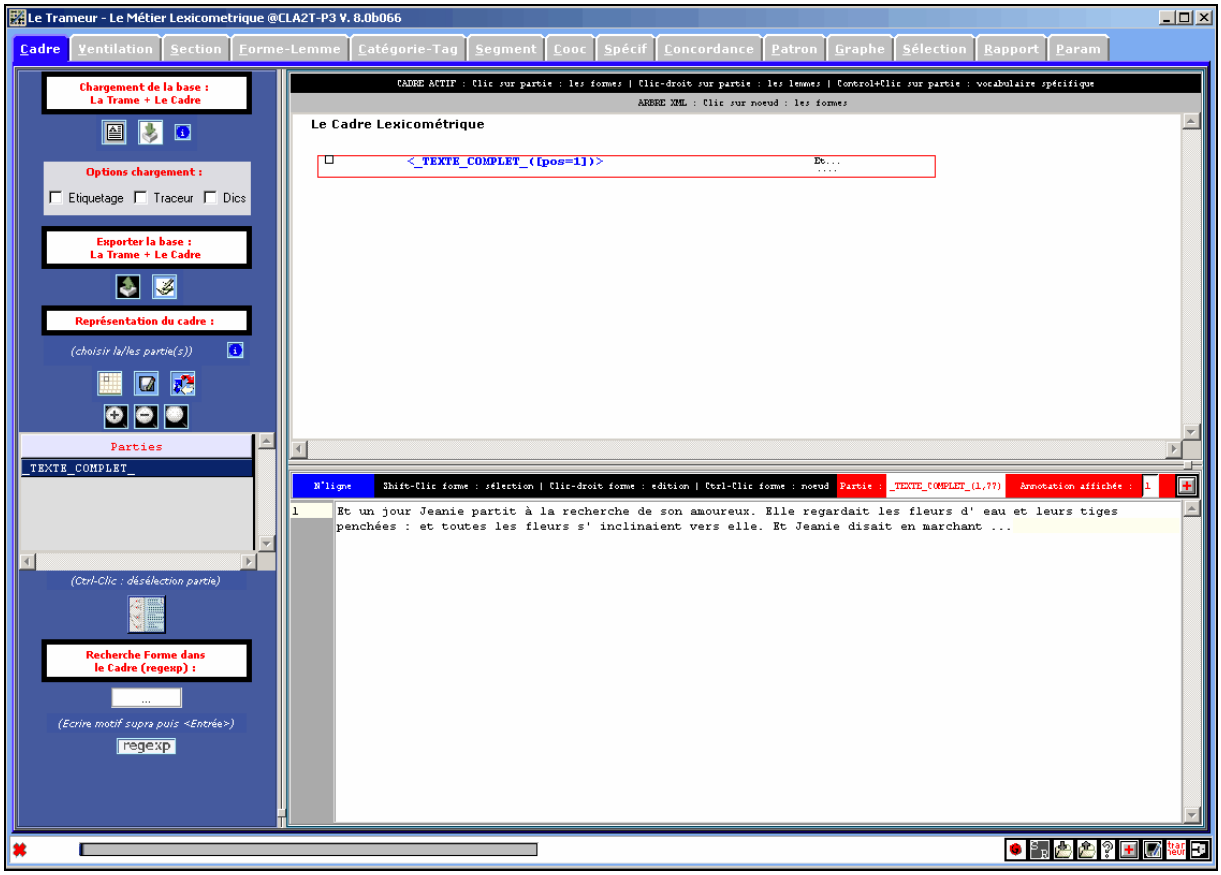


Figure 241 : Chargement de la base, accès à la *Trame* par le *Cadre*

L'éditeur de la ***Trame*** (accès par le ***Cadre*** ici), permet de définir des nœuds : le raccourci clavier *Control-Clic* sur un item de la ***Trame*** permet de le sélectionner, dans l'exemple suivant, les 4 premiers items sont sélectionnés, ils vont être utilisés pour définir le premier nœud du graphe.

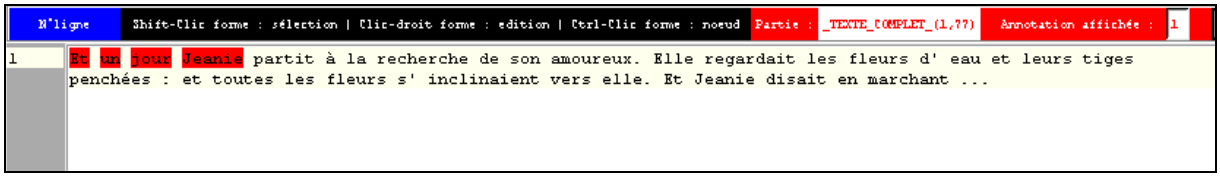

N°ligne
Shift-Clic forme : sélection | Clic-droit forme : edition | Ctrl-Clic forme : noeud
Partie : _TEXTE_COMPLET_(1,77)
Annotation affichée : 1
1
Et un jour Jeanie partit à la recherche de son amoureux. Elle regardait les fleurs d' eau et leurs tiges
penchées : et toutes les fleurs s' inclinaient vers elle. Et Jeanie disait en marchant ...

Figure 242 : Sélection des items constituant un nœud du graphe

On peut ensuite générer le nœud (`nœud1`) (*double-clic* dans l'éditeur) et le donner à voir dans l'onglet **GRAPHE** :

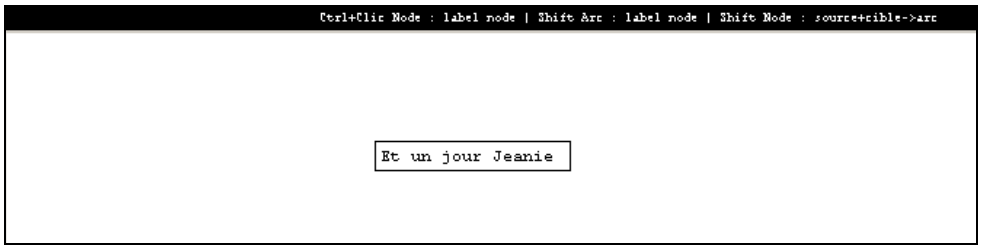


Figure 243 : Construction du nœud

Control-Clic sur le nœud permet de lui associer un label (valeur qui sera associée à chaque item constituant le nœud défini pour l'annotation en cours de construction) :

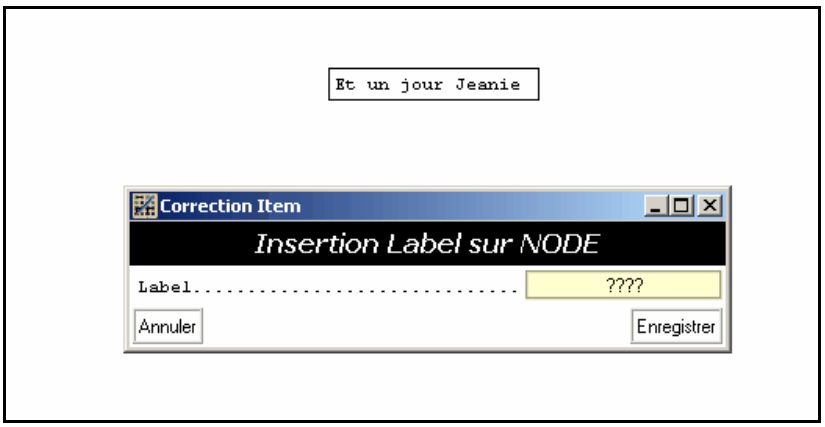


Figure 244 : Insertion d'un label sur le nœud

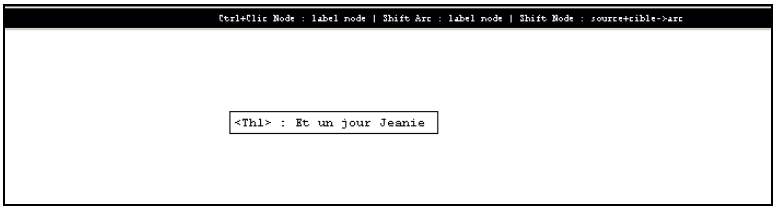


Figure 245 : Le nœud a un label

On peut ensuite réitérer l'opération de création d'un nœud (`nœud2`) :

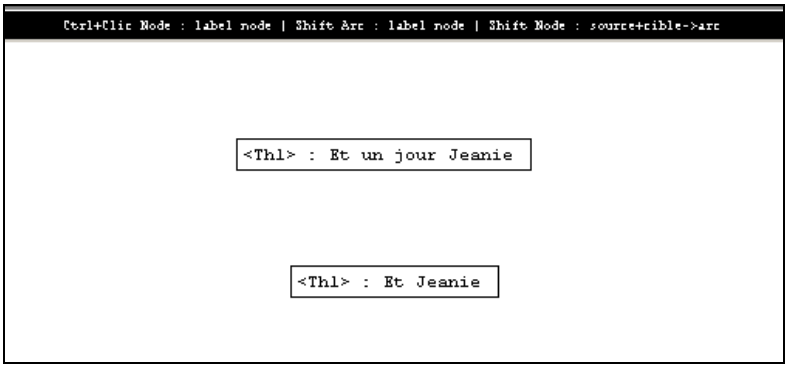


Figure 246 : De nouveaux nœuds…

Pour créer un arc entre `nœud1` et `nœud2`, on sélectionne le nœud source (*clic-droit* sur `nœud1` qui devient rouge) puis on sélectionne le nœud cible (*clic-droit* sur `nœud2` qui devient bleu). L'arc est ensuite généré :

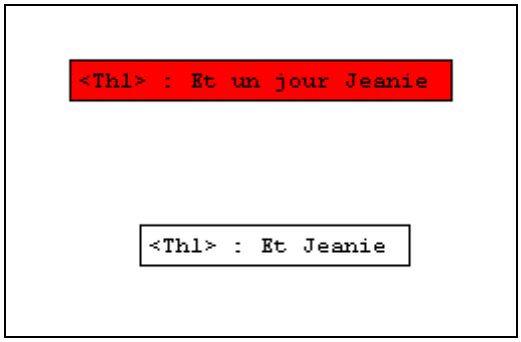


Figure 247 : La source d'un arc

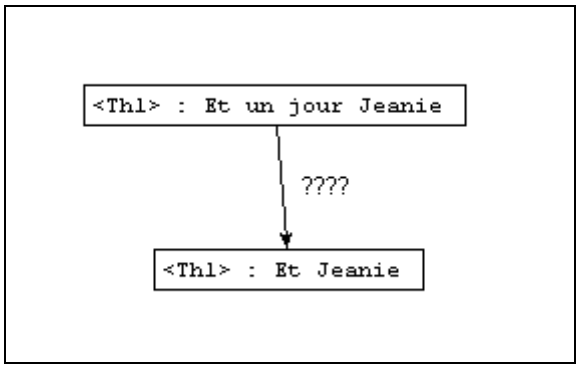


Figure 248 : La cible de l'arc... l'arc est défini

Control-Clic sur l'arc permet de lui associer un label :

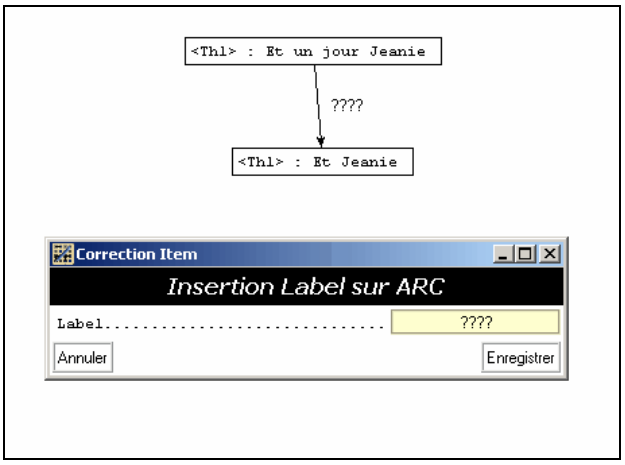


Figure 249 : Ajout d'un label sur l'arc

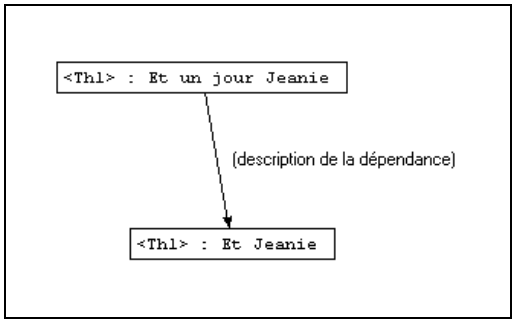


Figure 250 : L'arc a son label

Une fois le graphe complet construit,

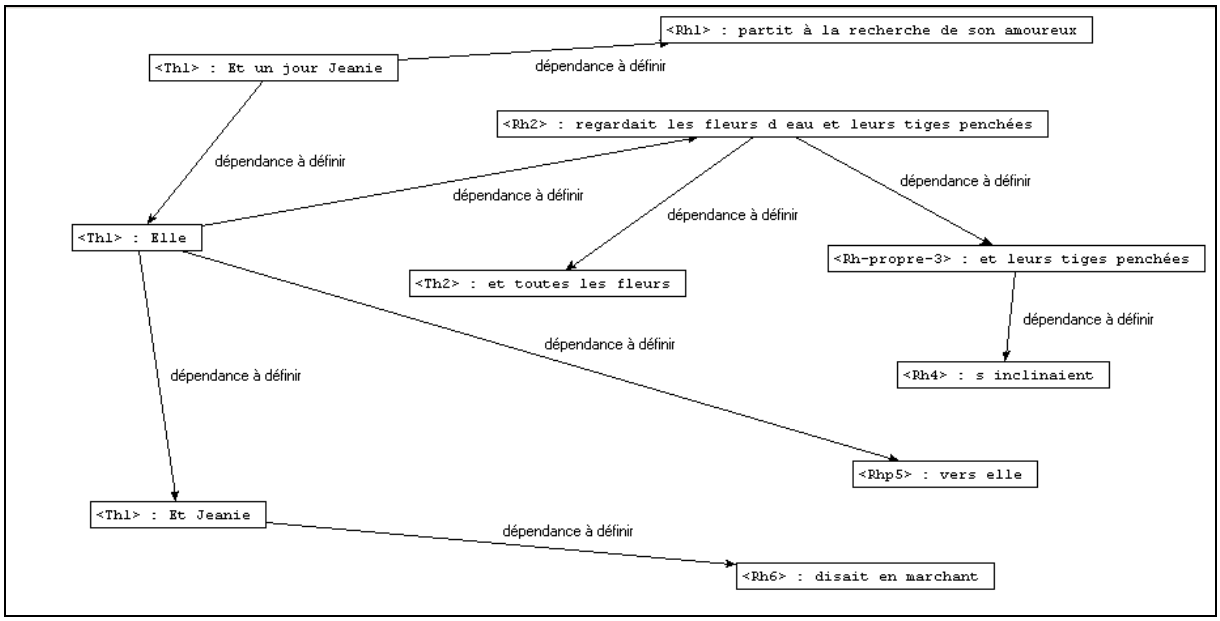


Figure 251 : Graphe complet construit

il est possible de le « générer », ce processus permet de créer un fichier dont le format est similaire à celui présenté *supra* :

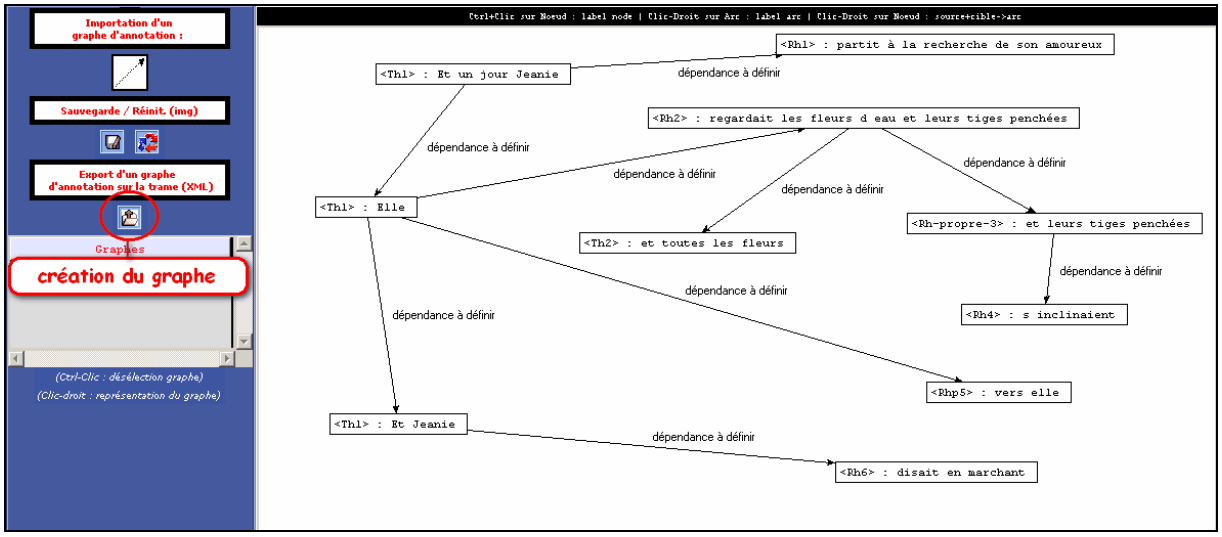

Importation d'un graphe d'annotation :
Sauvegarde / Réinit. (img)
Export d'un graphe d'annotation sur la trame (XML)
Graphes
création du graphe
(Ctrl-Clic : désélection graphe)
(Clic-droit : représentation du graphe)
Ctrl+Clic sur Noeud : label node | Clic-Droit sur Arc : label arc | Clic-Droit sur Noeud : source+cible->arc
<Rh1> : partit à la recherche de son amoureux
<Th1> : Et un jour Jeanie
dépendance à définir
<Rh2> : regardait les fleurs d eau et leurs tiges penchées
<Th1> : Elle
<Th2> : et toutes les fleurs
<Rh-propre-3> : et leurs tiges penchées
<Rh4> : s inclinaient
<Rhp5> : vers elle
<Th1> : Et Jeanie
<Rh6> : disait en marchant

Figure 252 : Sauvegarde du graphe

Ce fichier est ensuite ré-importable par le processus décrit dans la partie 1.

30.1.3 Exploitation des graphes d'annotation

Chantier en cours.

30.2 Base textométrique de textes alignés

30.2.1 Mise au jour d'alignement dans la carte des sections

Objectif : Mise au jour d'un alignement par sections dans la représentation graphique d'un corpus aligné via la *Carte des sections*

Corpus exemple : plusieurs traductions alignées du discours d'investiture de Barak Obama en 2009.

- **volet-en** : la transcription parue sur le site du New York Times
http://www.nytimes.com/2009/01/20/us/politics/20text-obama.html
- **volet -fr-1** : la traduction du discours mise en ligne sur le site du journal Le Monde (URL non accessible hors abonnement)
- **volet -fr-2** : la traduction du discours mise en ligne sur le site du journal Libération (AFP)
http://www.liberation.fr/monde/0101313402-point-par-point-le-discours-d-investiture-de-barack-obama-nouveau-president-des-etats-unis
- **volet -fr-3** : la traduction du discours mise en ligne sur le site de RFI
http://www.rfi.fr/actufr/articles/109/article_77584.asp
- **volet -fr-4** : la traduction du discours via Google de **volet-en**
http://translate.google.fr/

Ces 5 volets sont disponibles dans un fichier au format TMX.

On présente ci-dessous une illustration des différents paramétrages d'affichage de la carte des sections sur un corpus exemple décrit ci-dessus.

Méthode n°1 : Lecture du fichier alignement au format TMX

- Le ***Trameur*** peut lire directement le fichier au format TMX contenant les 5 volets
- Une fois chargé, on peut visualiser cet alignement *via* la carte des sections en choisissant comme délimiteur de partie l'élément **tu** (bloc de traduction contenant les différents volets de la même zone textuelle (phrases par exemple))
- **L'alignement à gauche** des sections permet de visualiser globalement tous les blocs et leurs différents volets (par section)

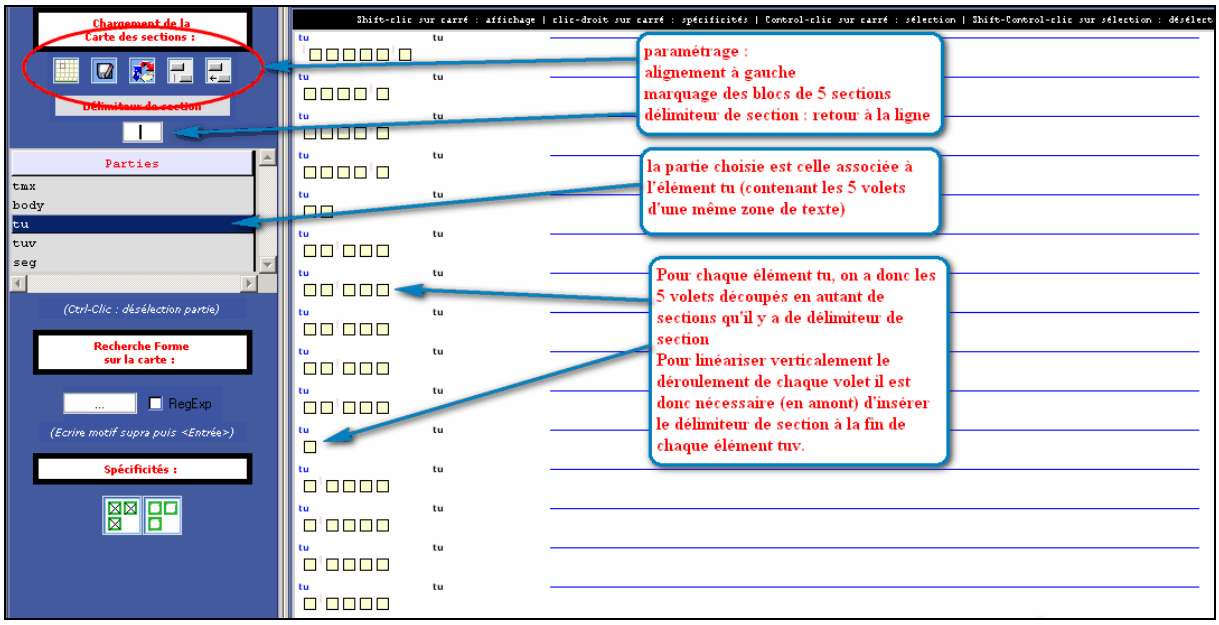

Chargement de la Carte des sections :
Délimiteur de section
Parties
tmx
body
tu
tuv
seg
(Ctrl-Clic : désélection partie)
Recherche Forme sur la carte :
RegExp
(Ecrire motif supra puis <Entrée>)
Spécificités :
Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : déselect
paramétrage :
alignement à gauche
marquage des blocs de 5 sections
délimiteur de section : retour à la ligne
la partie choisie est celle associée à l'élément tu (contenant les 5 volets d'une même zone de texte)
Pour chaque élément tu, on a donc les 5 volets découpés en autant de sections qu'il y a de délimiteur de section
Pour linéariser verticalement le déroulement de chaque volet il est donc nécessaire (en amont) d'insérer le délimiteur de section à la fin de chaque élément tuv.

Figure 253 : Alignement, partition tu (1)

Le marquage de blocs de sections est ici maintenu.

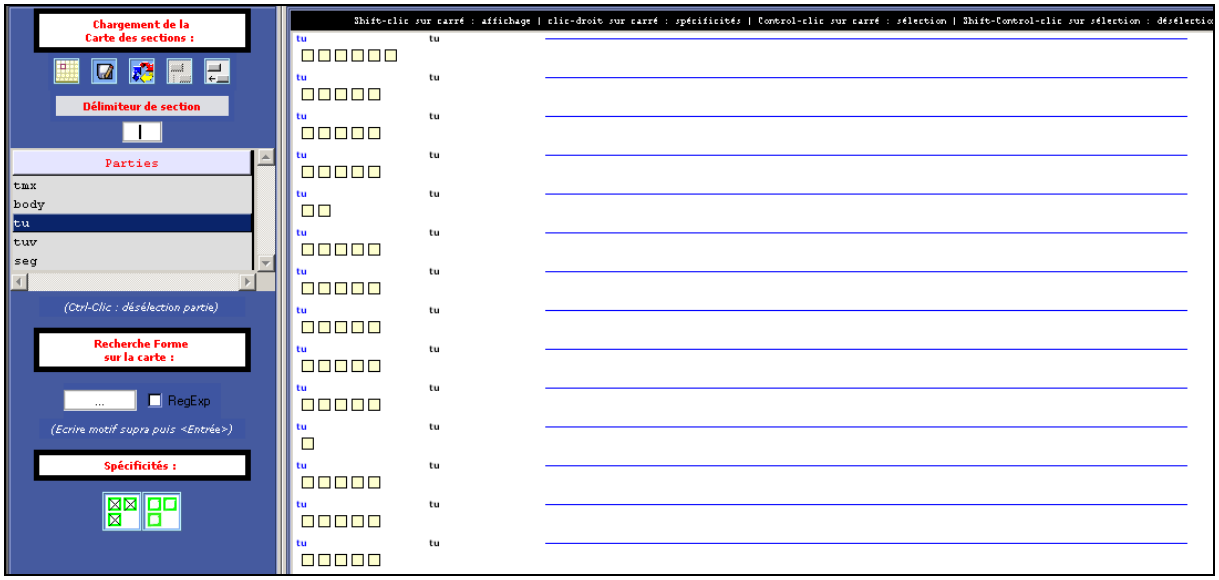

Chargement de la Carte des sections :
Délimiteur de section
Parties
tmx
body
tu
tuv
seg
(Ctrl-Clic : désélection partie)
Recherche Forme sur la carte :
RegExp
(Ecrire motif supra puis <Entrée>)
Spécificités :
Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : déselectio

Figure 254 : Alignement, partition tu (2)

Le marquage de blocs de sections est ici inhibé.

Inconvénient : on n'a pas la vue globale permettant de contraster les 4 volets

Méthode n°2 : Lecture du fichier alignement au format Lexico/*Trameur*

- Reconstruction des 4 volets :
  - A partir du TMX précédent, chargement du TMX dans ***mkAlign***[24]
  - sauvegarde de chacun des volets (format txt) : les éléments tuv du fichier TMX étant potentiellement une section
- Concaténation des 4 volets dans un même fichier et insertion de balise : <volet=en>...<volet=fr1> etc.
- Chargement dans le ***Trameur*** + carte des sections + **alignement à gauche**

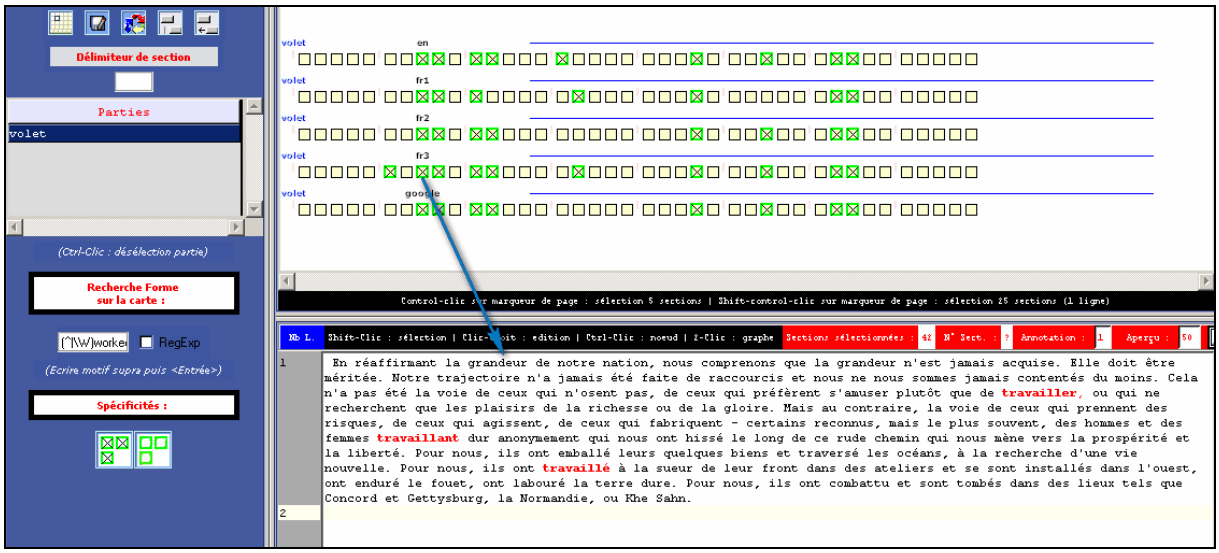

Délimiteur de section
Parties
volet
(Ctrl-Clic : désélection partie)
Recherche Forme sur la carte :
RegExp
(Ecrire motif supra puis <Entrée>)
Spécificités :
volet en
volet fr1
volet fr2
volet fr3
volet google
En réaffirmant la grandeur de notre nation, nous comprenons que la grandeur n'est jamais acquise. Elle doit être méritée. Notre trajectoire n'a jamais été faite de raccourcis et nous ne nous sommes jamais contentés du moins. Cela n'a pas été la voie de ceux qui n'osent pas, de ceux qui préfèrent s'amuser plutôt que de travailler, ou qui ne recherchent que les plaisirs de la richesse ou de la gloire. Mais au contraire, la voie de ceux qui prennent des risques, de ceux qui agissent, de ceux qui fabriquent - certains reconnus, mais le plus souvent, des hommes et des femmes travaillant dur anonymement qui nous ont hissé le long de ce rude chemin qui nous mène vers la prospérité et la liberté. Pour nous, ils ont emballé leurs quelques biens et traversé les océans, à la recherche d'une vie nouvelle. Pour nous, ils ont travaillé à la sueur de leur front dans des ateliers et se sont installés dans l'ouest, ont enduré le fouet, ont labouré la terre dure. Pour nous, ils ont combattu et sont tombés dans des lieux tels que Concord et Gettysburg, la Normandie, ou Khe Sahn.

Figure 255 : Alignement, partition volet (1)

Marquage des blocs de sections (projection de la famille de mots « travail / work » sur la carte des sections).

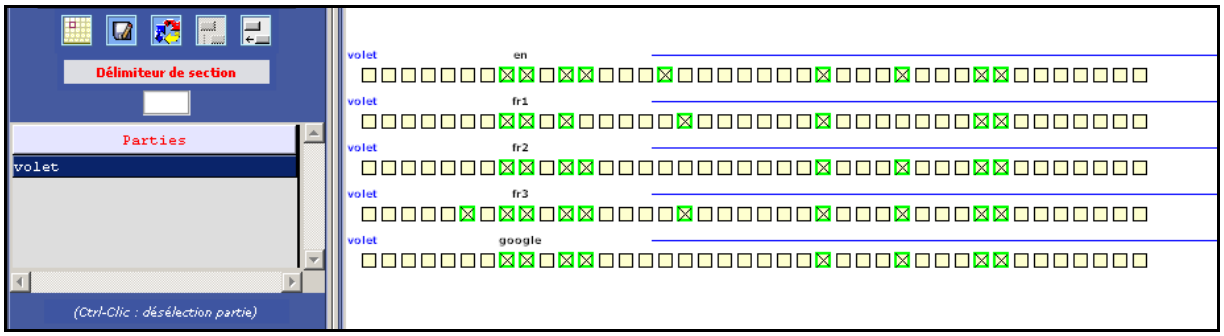


Figure 256 : Alignement, partition volet (2)

Inhibition du marquage des blocs de sections.

Ici, les 4 volets alignés sont visualisables globalement (par section (tuv dans le TMX de départ))

[24] http://tal.univ-paris3.fr/mkAlign/

30.2.2 Exploration d'une base textométrique de textes alignés

Cette section présente les fonctionnalités permettant de d'explorer une base intégrant 2 (ou plusieurs) textes alignés. Cette mise à jour reprend une fonctionnalité déjà disponible dans *mkAlign*[25] permettant notamment de mettre à jour des équivalents traductionnels dans des textes alignés par un calcul de cooccurrence réalisé en parallèle sur 2 volets de l'alignement traité.

30.2.2.1 Chargement d'une nouvelle base intégrant 2 textes bilingues alignés

Pour illustrer cette fonctionnalité, nous commençons par choisir un exemple de texte bilingue à charger par la suite dans le ***Trameur***. Nous travaillerons ici sur un texte de convention européenne disponible en français et en anglais, ces 2 volets étant alignés (ces 2 fichiers sont disponibles dans le programme *mkAlign*).

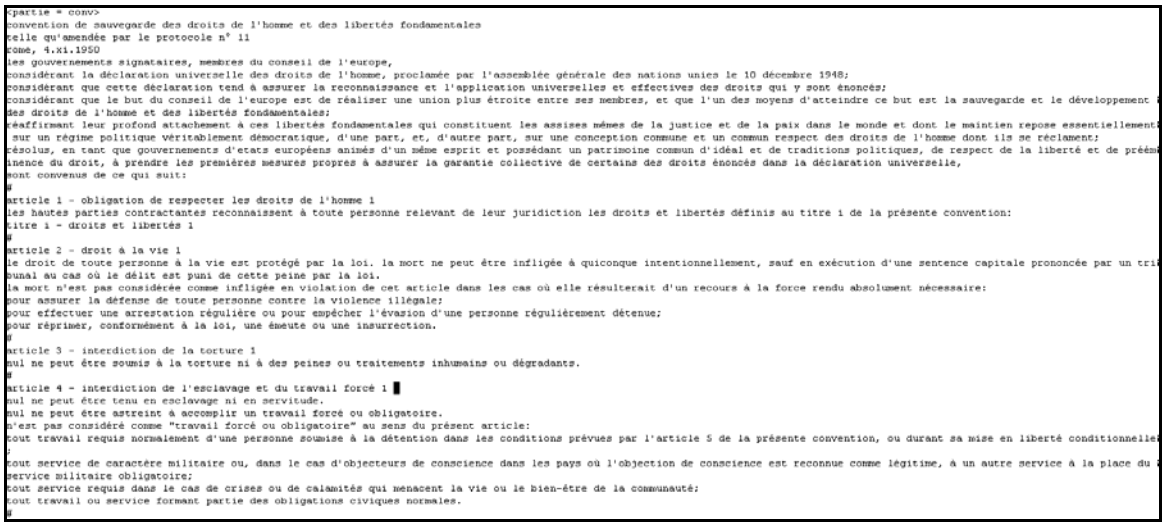

<partie = conv>
convention de sauvegarde des droits de l'homme et des libertés fondamentales
telle qu'amendée par le protocole n° 11
rome, 4.xi.1950
les gouvernements signataires, membres du conseil de l'europe,
considérant la déclaration universelle des droits de l'homme, proclamée par l'assemblée générale des nations unies le 10 décembre 1948;
considérant que cette déclaration tend à assurer la reconnaissance et l'application universelles et effectives des droits qui y sont énoncés;
considérant que le but du conseil de l'europe est de réaliser une union plus étroite entre ses membres, et que l'un des moyens d'atteindre ce but est la sauvegarde et le développement
des droits de l'homme et des libertés fondamentales;
réaffirmant leur profond attachement à ces libertés fondamentales qui constituent les assises mêmes de la justice et de la paix dans le monde et dont le maintien repose essentiellement
sur un régime politique véritablement démocratique, d'une part, et, d'autre part, sur une conception commune et un commun respect des droits de l'homme dont ils se réclament;
résolus, en tant que gouvernements d'etats européens animés d'un même esprit et possédant un patrimoine commun d'idéal et de traditions politiques, de respect de la liberté et de préém
inence du droit, à prendre les premières mesures propres à assurer la garantie collective de certains des droits énoncés dans la déclaration universelle,
sont convenus de ce qui suit:
#
article 1 - obligation de respecter les droits de l'homme 1
les hautes parties contractantes reconnaissent à toute personne relevant de leur juridiction les droits et libertés définis au titre i de la présente convention:
titre i - droits et libertés 1
#
article 2 - droit à la vie 1
le droit de toute personne à la vie est protégé par la loi. la mort ne peut être infligée à quiconque intentionnellement, sauf en exécution d'une sentence capitale prononcée par un tri
bunal au cas où le délit est puni de cette peine par la loi.
la mort n'est pas considérée comme infligée en violation de cet article dans les cas où elle résulterait d'un recours à la force rendu absolument nécessaire:
pour assurer la défense de toute personne contre la violence illégale;
pour effectuer une arrestation régulière ou pour empêcher l'évasion d'une personne régulièrement détenue;
pour réprimer, conformément à la loi, une émeute ou une insurrection.
#
article 3 - interdiction de la torture 1
nul ne peut être soumis à la torture ni à des peines ou traitements inhumains ou dégradants.
#
article 4 - interdiction de l'esclavage et du travail forcé 1
nul ne peut être tenu en esclavage ni en servitude.
nul ne peut être astreint à accomplir un travail forcé ou obligatoire.
n'est pas considéré comme "travail forcé ou obligatoire" au sens du présent article:
tout travail requis normalement d'une personne soumise à la détention dans les conditions prévues par l'article 5 de la présente convention, ou durant sa mise en liberté conditionnelle
;
tout service de caractère militaire ou, dans le cas d'objecteurs de conscience dans les pays où l'objection de conscience est reconnue comme légitime, à un autre service à la place du
service militaire obligatoire;
tout service requis dans le cas de crises ou de calamités qui menacent la vie ou le bien-être de la communauté;
tout travail ou service formant partie des obligations civiques normales.
#

Figure 257 : fichier 1, volet français (extrait)

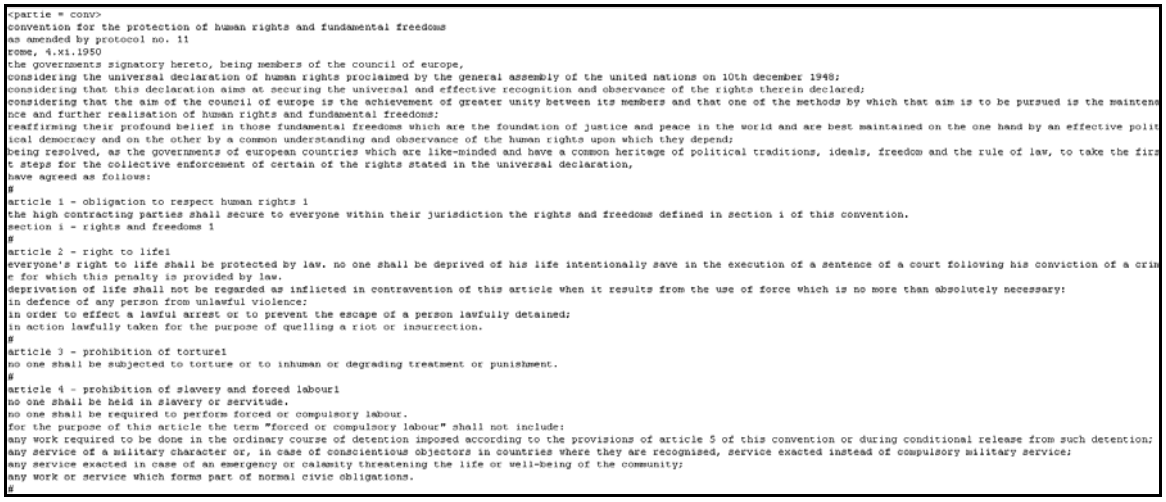

<partie = conv>
convention for the protection of human rights and fundamental freedoms
as amended by protocol no. 11
rome, 4.xi.1950
the governments signatory hereto, being members of the council of europe,
considering the universal declaration of human rights proclaimed by the general assembly of the united nations on 10th december 1948;
considering that this declaration aims at securing the universal and effective recognition and observance of the rights therein declared;
considering that the aim of the council of europe is the achievement of greater unity between its members and that one of the methods by which that aim is to be pursued is the maintena
nce and further realisation of human rights and fundamental freedoms;
reaffirming their profound belief in those fundamental freedoms which are the foundation of justice and peace in the world and are best maintained on the one hand by an effective polit
ical democracy and on the other by a common understanding and observance of the human rights upon which they depend;
being resolved, as the governments of european countries which are like-minded and have a common heritage of political traditions, ideals, freedom and the rule of law, to take the firs
t steps for the collective enforcement of certain of the rights stated in the universal declaration,
have agreed as follows:
#
article 1 - obligation to respect human rights 1
the high contracting parties shall secure to everyone within their jurisdiction the rights and freedoms defined in section i of this convention.
section i - rights and freedoms 1
#
article 2 - right to life1
everyone's right to life shall be protected by law. no one shall be deprived of his life intentionally save in the execution of a sentence of a court following his conviction of a crim
e for which this penalty is provided by law.
deprivation of life shall not be regarded as inflicted in contravention of this article when it results from the use of force which is no more than absolutely necessary:
in defence of any person from unlawful violence;
in order to effect a lawful arrest or to prevent the escape of a person lawfully detained;
in action lawfully taken for the purpose of quelling a riot or insurrection.
#
article 3 - prohibition of torture1
no one shall be subjected to torture or to inhuman or degrading treatment or punishment.
#
article 4 - prohibition of slavery and forced labour1
no one shall be held in slavery or servitude.
no one shall be required to perform forced or compulsory labour.
for the purpose of this article the term "forced or compulsory labour" shall not include:
any work required to be done in the ordinary course of detention imposed according to the provisions of article 5 of this convention or during conditional release from such detention;
any service of a military character or, in case of conscientious objectors in countries where they are recognised, service exacted instead of compulsory military service;
any service exacted in case of an emergency or calamity threatening the life or well-being of the community;
any work or service which forms part of normal civic obligations.
#

Figure 258 : fichier 1, volet anglais (extrait)

Les 2 textes sont structurés en partie (*via* un système de balisage). L'alignement est matérialisé dans les 2 volets par le caractère # qui induit les alignements respectifs des différents paragraphes.

[25] http://www.tal.univ-paris3.fr/mkAlign/

Dans *mkAlign*, les 2 fichiers sont chargés un à un (le volet FR et le volet EN) dans les 2 volets de l'alignement.

<partie = conv>
convention de sauvegarde des droits de l'homme et des libertés fondamentales
telle qu'amendée par le protocole n° 11
rome, 4.xi.1950
<partie = conv>
convention for the protection of human rights and fundamental freedoms
as amended by protocol no. 11
rome, 4.xi.1950
article 1 - obligation de respecter les droits de l'homme 1
les hautes parties contractantes reconnaissent à toute personne relevant de leur juridiction les droits et libertés définis au titre i de la présente convention:
article 1 - obligation to respect human rights 1
the high contracting parties shall secure to everyone within their jurisdiction the rights and freedoms defined in section i of this convention.
article 2 - droit à la vie 1
le droit de toute personne à la vie est protégé par la loi. la mort ne peut être infligée à quiconque intentionnellement, sauf en exécution d'une sentence capitale prononcée par un tribunal au cas où le délit
article 2 - right to life1
everyone's right to life shall be protected by law. no one shall be deprived of his life intentionally save in the execution of a sentence of a court following his conviction of a crime for which this
article 3 - interdiction de la torture 1
nul ne peut être soumis à la torture ni à des peines ou traitements inhumains ou dégradants.
article 3 - prohibition of torturel
no one shall be subjected to torture or to inhuman or degrading treatment or punishment.
article 4 - interdiction de l'esclavage et du travail forcé 1
nul ne peut être tenu en esclavage ni en servitude.
nul ne peut être astreint à accomplir un travail forcé ou obligatoire.
article 4 - prohibition of slavery and forced labourl
no one shall be held in slavery or servitude.
no one shall be required to perform forced or compulsory labour.
for the purpose of this article the term "forced or

Figure 259 : Edition d'un alignement dans mkAlign

Pour le ***Trameur***, la base textométrique est construite ici par concaténation des 2 fichiers en un seul fichier, en intégrant en outre un marqueur de partie supplémentaire pour contraster les 2 volets (`<volet="FR">` et `<volet="EN">` au début de chacun des 2 volets).

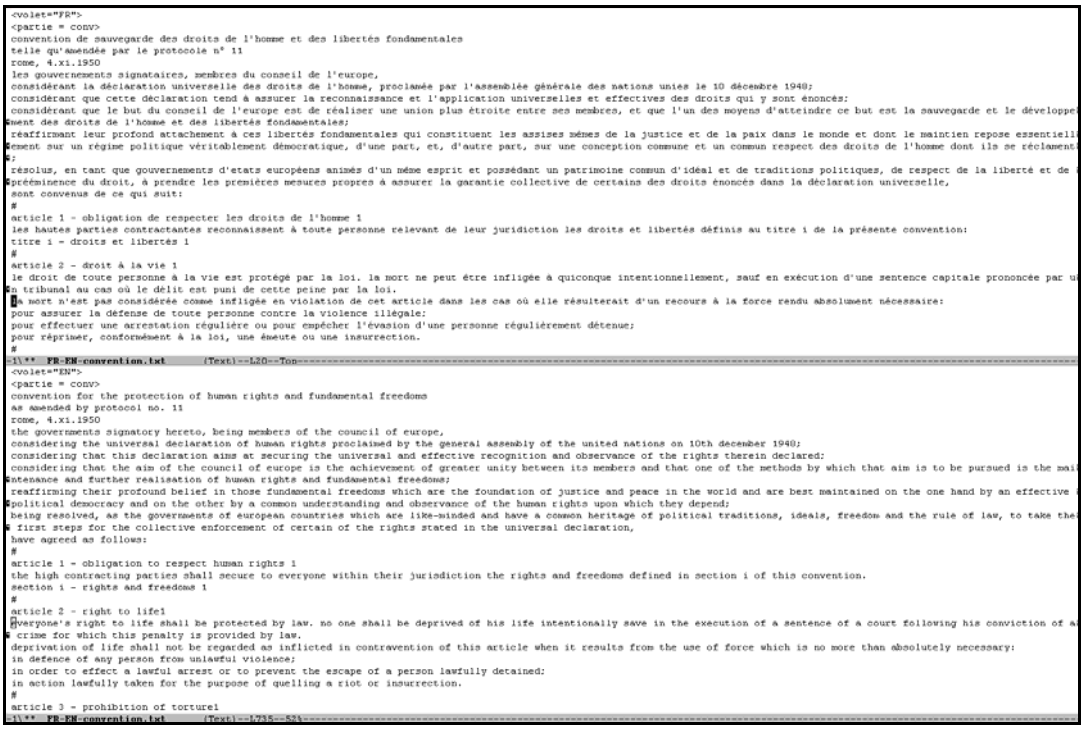

<volet="FR">
<partie = conv>
convention de sauvegarde des droits de l'homme et des libertés fondamentales
telle qu'amendée par le protocole n° 11
rome, 4.xi.1950
les gouvernements signataires, membres du conseil de l'europe,
considérant la déclaration universelle des droits de l'homme, proclamée par l'assemblée générale des nations unies le 10 décembre 1948;
considérant que cette déclaration tend à assurer la reconnaissance et l'application universelles et effectives des droits qui y sont énoncés;
considérant que le but du conseil de l'europe est de réaliser une union plus étroite entre ses membres, et que l'un des moyens d'atteindre ce but est la sauvegarde et le développe
ment des droits de l'homme et des libertés fondamentales;
réaffirmant leur profond attachement à ces libertés fondamentales qui constituent les assises mêmes de la justice et de la paix dans le monde et dont le maintien repose essentiell
ement sur un régime politique véritablement démocratique, d'une part, et, d'autre part, sur une conception commune et un commun respect des droits de l'homme dont ils se réclament
;
résolus, en tant que gouvernements d'etats européens animés d'un même esprit et possédant un patrimoine commun d'idéal et de traditions politiques, de respect de la liberté et de
prééminence du droit, à prendre les premières mesures propres à assurer la garantie collective de certains des droits énoncés dans la déclaration universelle,
sont convenus de ce qui suit:
#
article 1 - obligation de respecter les droits de l'homme 1
les hautes parties contractantes reconnaissent à toute personne relevant de leur juridiction les droits et libertés définis au titre i de la présente convention:
titre i - droits et libertés 1
#
article 2 - droit à la vie 1
le droit de toute personne à la vie est protégé par la loi. la mort ne peut être infligée à quiconque intentionnellement, sauf en exécution d'une sentence capitale prononcée par u
n tribunal au cas où le délit est puni de cette peine par la loi.
la mort n'est pas considérée comme infligée en violation de cet article dans les cas où elle résulterait d'un recours à la force rendu absolument nécessaire:
pour assurer la défense de toute personne contre la violence illégale;
pour effectuer une arrestation régulière ou pour empêcher l'évasion d'une personne régulièrement détenue;
pour réprimer, conformément à la loi, une émeute ou une insurrection.
#
-1** FR-EN-convention.txt (Text)--L20--Ton
<volet="EN">
<partie = conv>
convention for the protection of human rights and fundamental freedoms
as amended by protocol no. 11
rome, 4.xi.1950
the governments signatory hereto, being members of the council of europe,
considering the universal declaration of human rights proclaimed by the general assembly of the united nations on 10th december 1948;
considering that this declaration aims at securing the universal and effective recognition and observance of the rights therein declared;
considering that the aim of the council of europe is the achievement of greater unity between its members and that one of the methods by which that aim is to be pursued is the mai
ntenance and further realisation of human rights and fundamental freedoms;
reaffirming their profound belief in those fundamental freedoms which are the foundation of justice and peace in the world and are best maintained on the one hand by an effective
political democracy and on the other by a common understanding and observance of the human rights upon which they depend;
being resolved, as the governments of european countries which are like-minded and have a common heritage of political traditions, ideals, freedom and the rule of law, to take the
first steps for the collective enforcement of certain of the rights stated in the universal declaration,
have agreed as follows:
#
article 1 - obligation to respect human rights 1
the high contracting parties shall secure to everyone within their jurisdiction the rights and freedoms defined in section i of this convention.
section i - rights and freedoms 1
#
article 2 - right to life1
everyone's right to life shall be protected by law. no one shall be deprived of his life intentionally save in the execution of a sentence of a court following his conviction of a
crime for which this penalty is provided by law.
deprivation of life shall not be regarded as inflicted in contravention of this article when it results from the use of force which is no more than absolutely necessary:
in defence of any person from unlawful violence;
in order to effect a lawful arrest or to prevent the escape of a person lawfully detained;
in action lawfully taken for the purpose of quelling a riot or insurrection.
#
article 3 - prohibition of torturel
-1** FR-EN-convention.txt (Text)--L735--52%

Figure 260 : Préparation d'une base textométrique de textes bilingues alignés

La figure précédente donne à voir le fichier construit en mettant au jour les 2 volets (2 zones différentes du même fichier : le volet FR suivi du volet EN). Une fois chargée dans le ***Trameur*** (le caractère # ayant été ajouté à la liste des délimiteurs), la *Carte des sections* a l'allure suivante (avec le caractère # comme délimiteur de sections):

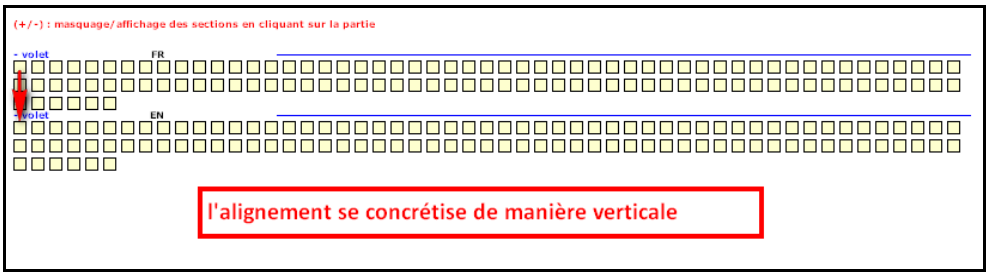

(+/-) : masquage/affichage des sections en cliquant sur la partie
- volet FR
- volet EN
l'alignement se concrétise de manière verticale

Figure 261 : Carte des sections d'un alignement dans le Trameur

On choisit d'afficher la partition VOLET pour contraster les 2 volets et on aligne à gauche les sections. Dans *mkAlign*, la *Carte des sections* de l'alignement a l'allure suivante :

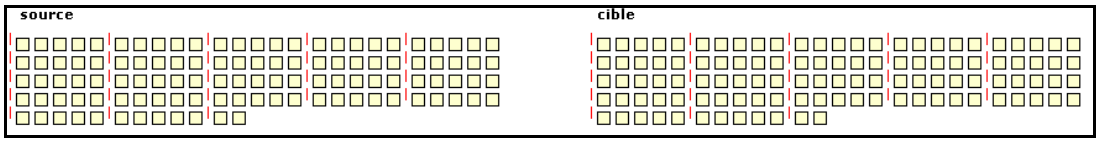


Figure 262 : Carte des sections d'un alignement dans mkAlign

L'alignement se lit de manière horizontale. Contrairement à *mkAlign*, l'alignement produit *via* la *Carte des sections* dans le ***Trameur*** se lit de manière verticale ; dans la figure précédente, le premier carré FR a son "miroir" au « bout de la flèche »… Si on projette une forme sur la carte, on obtient ceci :

Dans *mkAlign* :

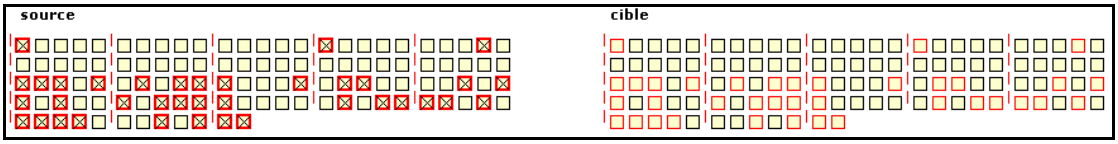


Figure 263 : Projection d'une forme sur carte des sections dans mkAlign

Le mot « conseil » est projeté sur le volet FR (à gauche), la zone miroir est matérialisée dans le volet EN par des carrés cerclés de rouge.

Dans le ***Trameur*** :

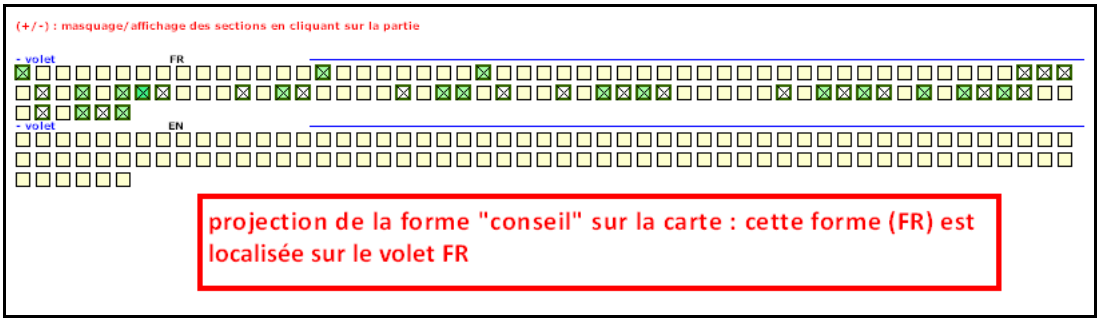

(+/-) : masquage/affichage des sections en cliquant sur la partie
- volet FR
- volet EN
projection de la forme "conseil" sur la carte : cette forme (FR) est localisée sur le volet FR

Figure 264 : Projection d'une forme sur la carte des sections dans Le Trameur

Le mot « conseil » est projeté sur la carte, seul le volet FR (en haut) met au jour la présence de la forme visée dans les sections cochées, la zone miroir n'est pas directement matérialisée dans le volet EN.
Dans le ***Trameur***, le calcul des spécificités des sections contenant une forme projetée tient compte du vocabulaire de l'ensemble du corpus (ici les 2 volets). Alors que dans *mkAlign*, ce calcul se fait sur le volet FR et en parallèle sur le volet EN (les 2 volets étant dissociés).
La nouvelle fonctionnalité introduite dans le ***Trameur*** consiste à "dissocier" les 2 volets de l'alignement et à reproduire un calcul en parallèle sur les 2 volets (*i.e* comme dans *mkAlign*) :

- Calcul des cooccurrences sur les sections contenant la forme projetée sur le volet FR
- En parallèle, calcul des cooccurrences sur les sections miroirs des précédentes

Pour cela, on dispose d'un nouveau menu dans l'onglet "Section" :

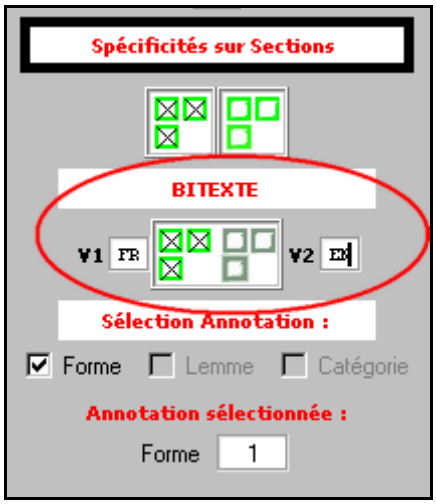


Figure 265 : Paramétrage d'un bitexte

Dans la zone BITEXTE : il est possible de considérer la *Carte des sections* comme une concaténation de 2 volets alignés.

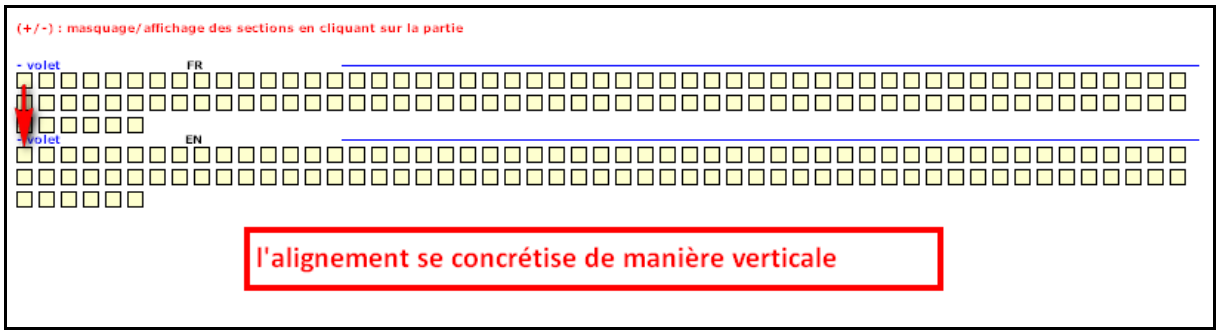


Figure 266 : Mise au jour du bitexte via la carte des sections

C'est le cas ci-dessus, pour la partition `VOLET`, on dispose bien d'une partie `VOLET=FR` et d'une partie `VOLET=EN`, les sections de chacun de ces volets étant alignées verticalement.
Pour amorcer le calcul, on paramètre donc les 2 volets à contraster : le volet V1 est celui contenant la forme projetée (ici FR est renseignée dans la zone de saisie V1) et le volet V2 est le miroir (ici EN dans la zone de saisie V2).
Le bouton du milieu permet de lancer le calcul qui produit ceci : (2 extraits : volet SOURCE et volet CIBLE)

Specificites du vocabulaire (Annotation Forme) sur les sections SOURCE contenant le motif : conseil

Nombre d'occurrences du texte global = 11099
Nombre d'occurrences dans la partie visee = 5314
Seuil : 10

---------- SOURCE ----------

---------- (cible infra) ----------

(Specificites positives en haut de liste, negatives en bas)
Le fichier construit : C:/SERGEF-1/SF-WORK/07-DEVEL/ANR-TE-1/MKGRAP-1/textes/CONVEN-1/FR-EN/export/resultspecif-source-1414344563.txt

Forme	Ind-Specif	Fq-Totale	Fq-Partie
conseil	26.4	79	79
europe	26.1	78	78
du	19.6	192	152
protocole	18.1	95	85
secrétaire	16.1	47	47
déclaration	15.0	49	48
ratification	13.4	44	43
le	12.9	231	163
général	12.6	50	47
membres	9.7	32	31
présent	8.9	83	65
vigueur	8.5	36	33
suit	8.4	23	23
acceptation	8.0	22	22
approbation	7.7	21	21
signature	7.6	25	24
mois	7.4	32	29
qui	7.3	91	67
territoires	7.1	19	19
premier	7.1	19	19
entrera	6.8	18	18
n°	6.8	18	18
jour	6.4	17	17
tout	6.1	67	50
notification	6.1	16	16
date	6.0	27	24

Figure 267 : Spécificités volet SOURCE de l'alignement

---------- CIBLE ----------

Specificites du vocabulaire (Annotation Forme) sur les sections CIBLE miroir de celles contenant le motif : conseil

Nombre d'occurrences du texte global = 10453
Nombre d'occurrences dans la partie visee = 5047
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)

Forme	Ind-Specif	Fq-Totale	Fq-Partie
council	25.2	76	76
europe	24.8	75	75
secretary	15.9	47	47
protocol	15.7	87	77
declaration	15.6	46	46
ratification	13.3	44	43
general	10.7	55	49
member	9.3	31	30
day	8.6	24	24
signature	8.6	24	24
into	8.4	28	27
of	8.1	873	496
acceptance	8.0	22	22
approval	7.6	21	21
date	7.3	32	29
at	6.7	54	43
first	6.3	21	20
signed	6.1	16	16
month	6.1	16	16
notification	6.1	16	16
force	6.1	31	27
territories	6.1	16	16
being	5.7	26	23
deposit	5.7	15	15
territory	5.1	33	27
enter	5.1	17	16
after	4.9	26	22
expiration	4.8	12	12
signatory	4.8	12	12
instrument	4.8	16	15
deposited	4.5	15	14
fundamental	4.4	18	16
rome	4.2	10	10
instruments	4.2	10	10
receipt	4.2	10	10
terms	4.2	10	10
addressed	4.2	10	10
this	4.0	118	74

Figure 268 : Spécificités volet CIBLE de l'alignement

Le calcul a produit le calcul des cooccurrences sur les sections du volet FR contenant la forme projetée (le mot « conseil »), *i.e* le volet SOURCE ; en parallèle ce calcul a aussi été réalisé sur le volet CIBLE : les sections miroir des précédentes dans lesquelles on trouve par

exemple l'équivalent traductionnel de « conseil » mis au jour par le calcul. C'est à dire la même chose que *via mkAlign* :

Specificites du vocabulaire sur les sections SOURCE contenant le motif : <conseil>

Nombre d'occurrences du texte global = 11099
Nombre d'occurrences dans la partie visee = 5314
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit : C:/SERGEF-1/SF-WORK/07-DEVEL/MKALIG-1/corpus/CONVEN-1/CONVEN-1/FR/export

Forme	Ind-Specif	Fq-Totale	Fq-Part
conseil	26.4	79	79
europe	26.1	78	78
du	19.6	192	152
protocole	18.1	95	85
secrétaire	16.1	47	47
déclaration	15.0	49	48
ratification	13.4	44	43
le	12.9	231	163
général	12.6	50	47
membres	9.7	32	31
présent	8.9	83	65
vigueur	8.5	36	33
suit	8.4	23	23
acceptation	8.0	22	22
approbation	7.7	21	21
signature	7.6	25	24
mois	7.4	32	29
qui	7.3	91	67
territoires	7.1	19	19
premier	7.1	19	19
entrera	6.8	18	18
n'	6.8	18	18
jour	6.4	17	17
tout	6.1	67	50

Specificites du vocabulaire sur les sections CIBLE associees aux sections SOURCE contenant

Nombre d'occurrences du texte global = 10453
Nombre d'occurrences dans la partie visee = 5047
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit : C:/SERGEF-1/SF-WORK/07-DEVEL/MKALIG-1/corpus/CONVEN-1/CONVEN-1/FR/export

Forme	Ind-Specif	Fq-Totale	Fq-Part
council	25.2	76	76
europe	24.8	75	75
secretary	15.9	47	47
protocol	15.7	87	77
declaration	15.6	46	46
ratification	13.3	44	43
general	10.7	55	49
member	9.3	31	30
day	8.6	24	24
signature	8.6	24	24
into	8.4	28	27
of	8.1	873	496
acceptance	8.0	22	22
approval	7.6	21	21
date	7.3	32	29
at	6.7	54	43
first	6.3	21	20
force	6.1	31	27
signed	6.1	16	16
month	6.1	16	16
notification	6.1	16	16
territories	6.1	16	16
deposit	5.7	15	15
being	5.7	26	23

Figure 269 : Spécificités des volets SOURCE et CIBLE via mkAlign

30.2.2.2 Chargement d'une nouvelle base intégrant 2 textes alignés

Le même type de calcul est possible sur une base intégrant 2 textes alignés d'une même langue, par exemple deux traductions alignées d'un même texte. Nous considérons ici 2 traductions[26] de l'ouvrage d'Hemingway « Le vieil homme et la mer » : celle de J. Dutourd (1952) et celle de F. Bon (2012).

```
<source="TRADUCTION1">
 Il était une fois un vieil homme, tout seul dans son bateau qui pêchait au milieu du Gulf Stream. En quatre-vin
gt-quatre jours, il n'avait pas pris un poisson. Les quarante premiers jours, un jeune garçon l'accompagna ; mai
s au bout de ce temps, les parents du jeune garçon déclarèrent que le vieux était décidément et sans remède sala
o ce qui veut dire aussi guignard qu'on peut l'être. On embarqua donc le gamin sur un autre bateau, lequel, en u
ne semaine, ramena trois poissons superbes. Chaque soir le gamin avait la tristesse de voir le vieux rentrer ave
c sa barque vide. Il ne manquait pas d'aller à sa rencontre et l'aidait à porter les lignes serrées en spirales,
 la gaffe, le harpon, ou la voile roulée autour du mât. La voile était rapiécée avec de vieux sacs de farine ; a
insi repliée, elle figurait le drapeau en berne de la défaite. #
 #
 #
 Le vieil homme était maigre et sec, avec des rides comme des coups de couteau sur la nuque. Les taches brunes d
e cet inoffensif cancer de la peau que cause la réverbération du soleil sur la mer des Tropiques marquaient ses
joues ; elles couvraient presque entièrement les deux côtés de son visage ; ses mains portaient les entailles pr
ofondes que font les filins au bout desquels se débattent les lourds poissons. Mais aucune de ces entailles n'ét
ait récente elles étaient vieilles comme les érosions d'un désert sans poissons. #
 #
 #
-u\--  VIEILHOMME-2-TRADUCTIONS.txt      (Text)--L1--Top-----------------------------------------------------------------
 <source="TRADUCTION2">
 Le vieil homme pêchait seul dans le Gulf Stream sur son canot depuis quatre-vingt-quatre jours sans avoir pris
un poisson. Les quarante premiers jours, le garçon était venu avec lui. Mais après ces quarante jours, les paren
ts du garçon lui avaient dit que le vieil homme était finalement et définitivement salao, ce qui est la pire for
me pour dire pas de chance, et selon leurs ordres, le garçon était parti sur un autre bateau, lequel avait pris
trois gros poissons la première semaine. Cela le rendait triste, le garçon, de voir le vieil homme revenir chaqu
e soir le canot vide, et toujours il le rejoignait pour l'aider à porter les lignes enroulées, la gaffe, le harp
on et la voile ferlée autour du mât. Une voile rapiécée avec des sacs de farine qui pendait ainsi comme le drape
au d'une permanente défaite. #
 #
 #
 Le vieil homme était maigre et hâve, avec de profondes rides dans l'arrière du cou. Sur ses joues, les taches b
runes d'un cancer de la peau bénin à cause de la réflexion du soleil sur la mer des tropiques. Les taches lui to
mbaient de chaque côté du visage et ses mains gardaient les cicatrices profondément plissées des poissons hâlés
lourdement sur la corde. Mais aucune de ces cicatrices pour être récente. Elles étaient aussi vieilles que l'éro
sion dans le désert sans poisson. #
 #
 #
 Tout en lui était vieux, sauf les yeux - et ils étaient de la même couleur que la mer, joyeux et invincibles. #
-u\--  VIEILHOMME-2-TRADUCTIONS.txt      (Text)--L653--51%----------------------------------------------------------------
```

Figure 270 : Base textométrique de 2 traductions alignées

Comme précédemment, une base regroupant les 2 traductions est obtenue en concaténant les 2 fichiers. La figure précédente donne à voir le début des 2 volets dans la base ainsi constituée. Une fois le fichier chargé, la *Carte des sections* a cette allure :

[26] http://www.tal.univ-paris3.fr/mkAlign/#p7b

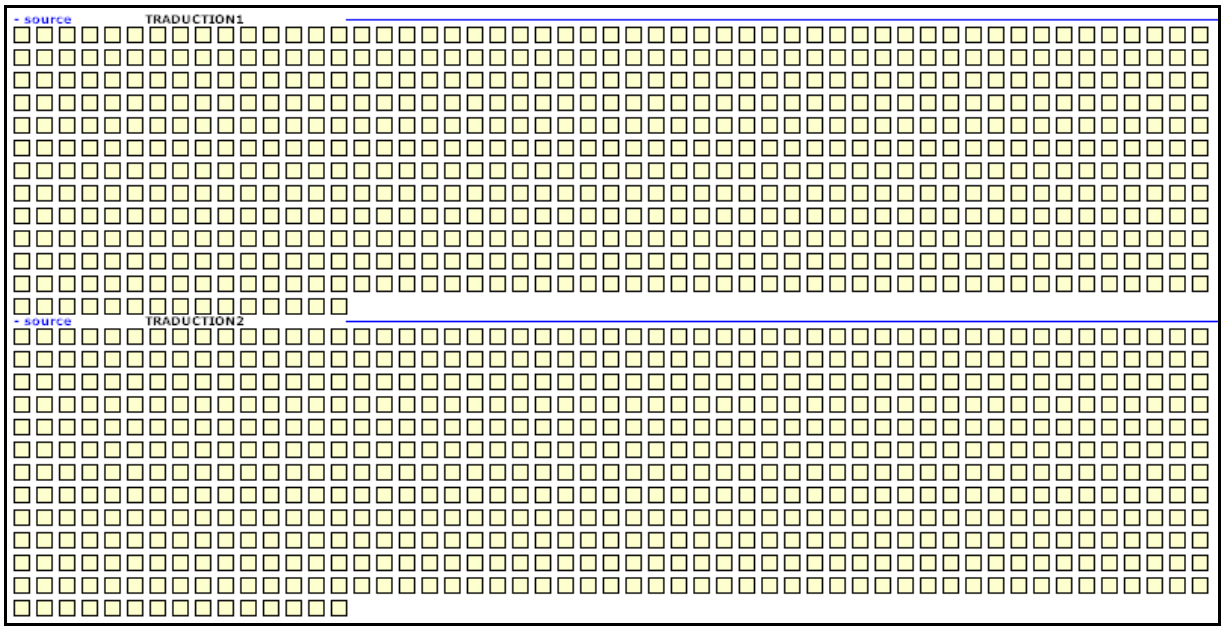


Figure 271 : Carte des sections d'un alignement de traductions dans le Trameur

Le volet « TRADUCTION1 » correspond à la traduction de Dutourd, le volet « TRADUCTION2 » à celle de Bon. La projection de la forme « vieil » sur la carte produit cette ventilation :

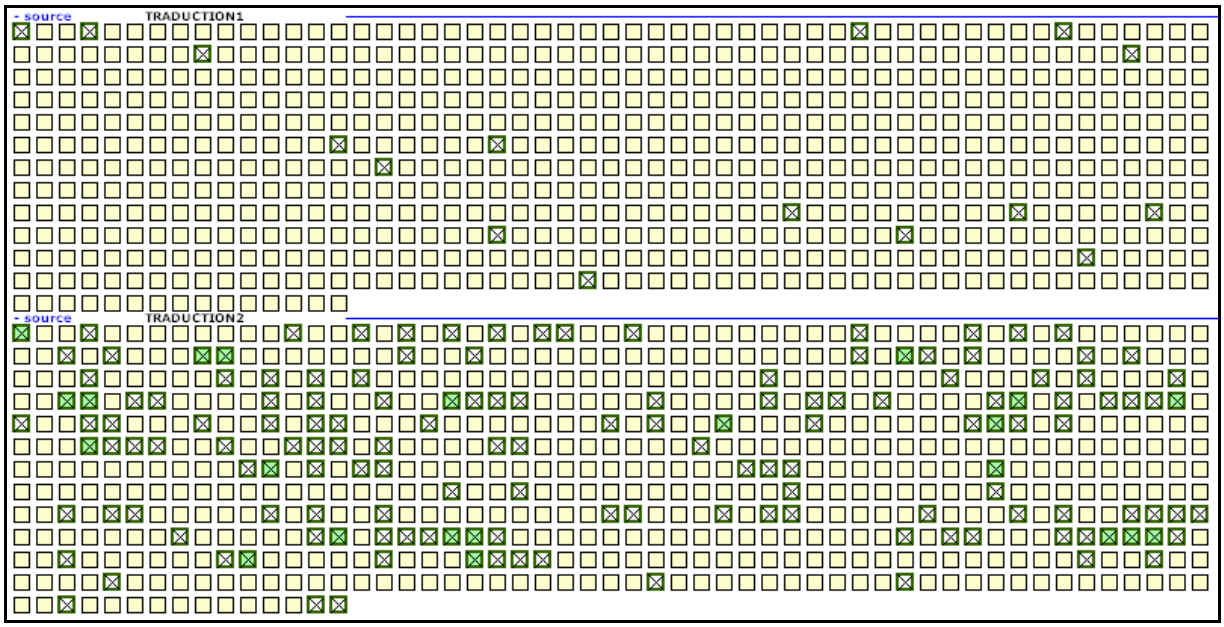


Figure 272 : Projection d'une forme sur les 2 traductions alignées

Cette forme est présente dans les 2 volets et massivement dans le second. Le calcul des cooccurrents présenté *supra* se fait en considérant les 2 volets comme 2 parties disjointes. Ici, la partie TRADUCTION2 est considérée comme le volet SOURCE et la partie TRADUCTION1 comme le volet CIBLE. Le résultat produit est le suivant :

Specificites du vocabulaire (Annotation Forme) sur les sections SOURCE contenant le motif : vieil

Nombre d'occurrences du texte global = 29257
Nombre d'occurrences dans la partie visee = 10191
Seuil : 10
---------- SOURCE ----------
---------- (cible infra) ----------
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit :
C:/DOCUME~1/FLEURY~1/MESDOC~1/MYDROP~1/WIP/WIP-TR~1/ALIGNE~1/ALIGNE~1/export/resultspecif-source-1414521180.txt

Forme	Ind-Specif	Eq-Totale	Eq-Partie
vieil	**	185	185
homme	**	211	189
Le	11.6	125	80
le	8.8	1005	434
requin	6.9	40	29
sur	4.7	175	84
ses	4.6	106	55
lame	4.3	10	9
était	4.3	251	113
étaient	4.0	40	24
yeux	3.9	28	18
nageoires	3.9	9	8
avaient	3.9	32	20
avait	3.8	177	81
parlaient	3.7	6	6
frappa	3.6	13	10
dents	3.5	8	7
continuait	3.3	10	8
Parfois	3.3	5	5
harpon	3.2	20	13
ainsi	3.1	7	6
surface	3.1	29	17
et	3.0	989	379
leurs	3.0	19	12
son	3.0	157	69
poissons	3.0	34	19
eau	2.9	73	35
chair	2.9	15	10
salée	2.8	4	4
élevait	2.8	4	4
cou	2.8	4	4

Figure 273 : Spécificités sur volet SOURCE (TRADUCTION2) : sections contenant le motif "vieil"

---------- CIBLE ----------
Specificites du vocabulaire (Annotation Forme) sur les sections CIBLE miroir de celles contenant le motif : vieil

Nombre d'occurrences du texte global = 29982
Nombre d'occurrences dans la partie visee = 10133
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)

Forme	Ind-Specif	Eq-Totale	Eq-Partie
vieux	32.8	269	186
Le	11.3	181	104
requin	7.7	35	27
vieil	7.0	16	15
le	5.6	889	358
avait	4.9	163	78
sa	4.7	123	61
la	4.5	826	326
du	4.1	209	93
lame	3.8	6	6
l	3.7	525	209
surface	3.7	15	11
eau	3.7	80	40
harpon	3.7	19	13
ses	3.6	98	47
étaient	3.6	43	24
Les	3.5	46	25
frappa	3.5	8	7
jaillissaient	3.4	5	5
chair	3.4	12	9
grosses	3.4	5	5
était	3.2	251	104
plusieurs	3.1	7	6
gueule	3.0	9	7
mâchoires	3.0	9	7
aperçut	3.0	17	11
côté	3.0	17	11
vit	3.0	11	8
laissait	2.9	4	4
faux	2.9	4	4
enfonça	2.9	4	4
poissons	2.9	33	18
pièce	2.9	4	4

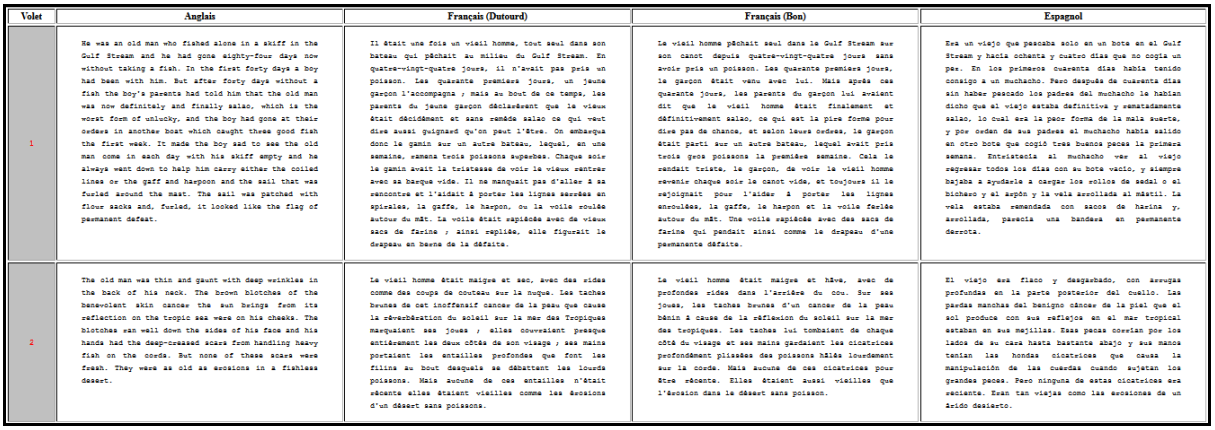
Figure 274 : Spécificités sur volet CIBLE (TRADUCTION1) :

sections MIROIR des sections contenant le motif "vieil"

Dans la traduction de F. Bon, on trouve plutôt la séquence « le vieil homme » et dans celle de J. Dutourd, la séquence « le vieux ».

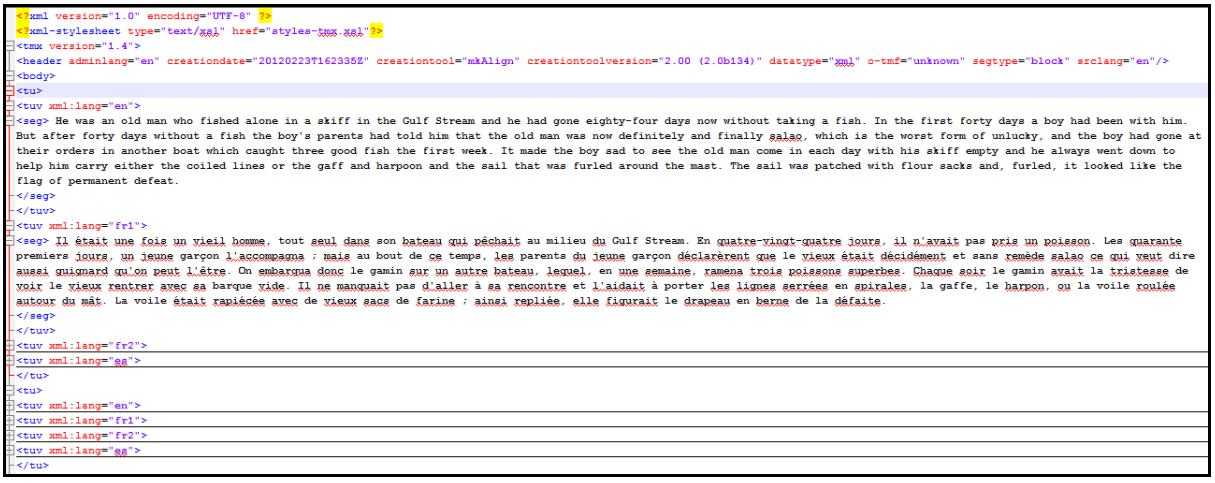
30.2.2.3 Base textométrique : alignement au format TMX

Un alignement au format TMX permet lui aussi de réaliser un calcul contrastif de cooccurrences sur 2 des volets intégrés dans ce type d'alignement. Nous prendrons ici comme exemple de fichier TMX, un alignement intégrant 4 volets de l'ouvrage “Le vieil homme et la mer” : le volet anglais original (noté en), les 2 volets traduits en français déjà mentionnés (noté fr1 et fr2) et un volet traduit en espagnol (noté es).
Le fichier a l'allure suivante (affichage partiel du fichier dans un navigateur, puis dans un éditeur (code source)):



Volet	Anglais	Français (Dutourd)	Français (Bon)	Espagnol
1	He was an old man who fished alone in a skiff in the Gulf Stream and he had gone eighty-four days now without taking a fish. In the first forty days a boy had been with him. But after forty days without a fish the boy's parents had told him that the old man was now definitely and finally salao, which is the worst form of unlucky, and the boy had gone at their orders in another boat which caught three good fish the first week. It made the boy sad to see the old man come in each day with his skiff empty and he always went down to help him carry either the coiled lines or the gaff and harpoon and the sail that was furled around the mast. The sail was patched with flour sacks and, furled, it looked like the flag of permanent defeat.	Il était une fois un vieil homme, tout seul dans son bateau qui pêchait au milieu du Gulf Stream. En quatre-vingt-quatre jours, il n'avait pas pris un poisson. Les quarante premiers jours, un jeune garçon l'accompagna ; mais au bout de ce temps, les parents du jeune garçon déclarèrent que le vieux était décidément et sans remède salao ce qui veut dire aussi guignard qu'on peut l'être. On embarqua donc le gamin sur un autre bateau, lequel, en une semaine, ramena trois poissons superbes. Chaque soir le gamin avait la tristesse de voir le vieux rentrer avec sa barque vide. Il ne manquait pas d'aller à sa rencontre et l'aidait à porter les lignes serrées en spirales, la gaffe, le harpon, ou la voile roulée autour du mât. La voile était rapiécée avec de vieux sacs de farine ; ainsi repliée, elle figurait le drapeau en berne de la défaite.	Le vieil homme pêchait seul dans le Gulf Stream sur son canot depuis quatre-vingt-quatre jours sans avoir pris un poisson. Les quarante premiers jours, le garçon était vêtu avec lui. Mais après ces quarante jours, les parents du garçon lui avaient dit que le vieil homme était finalement et définitivement salao, ce qui est la pire forme pour dire pas de chance, et selon leurs ordres, le garçon était parti sur un autre bateau, lequel avait pris trois gros poissons la première semaine. Cela le rendait triste, le garçon, de voir le vieil homme revenir chaque soir le canot vide, et toujours il le rejoignait pour l'aider à porter les lignes enroulées, la gaffe, le harpon et la voile ferlée autour du mât. Une voile rapiécée avec des sacs de farine qui pendait ainsi comme le drapeau d'une permanente défaite.	Era un viejo que pescaba solo en un bote en el Gulf Stream y hacía ochenta y cuatro días que no cogía un pez. En los primeros cuarenta días había tenido consigo a un muchacho. Pero después de cuarenta días sin haber pescado los padres del muchacho le habían dicho que el viejo estaba definitiva y rematadamente salao, lo cual era la peor forma de la mala suerte, y por orden de sus padres el muchacho había salido en otro bote que cogió tres buenos peces la primera semana. Entristecía al muchacho ver al viejo regresar todos los días con su bote vacío, y siempre bajaba a ayudarle a cargar los rollos de sedal o el bichero y el arpón y la vela arrollada al mástil. La vela estaba remendada con sacos de harina y, arrollada, parecía una bandera en permanente derrota.
2	The old man was thin and gaunt with deep wrinkles in the back of his neck. The brown blotches of the benevolent skin cancer the sun brings from its reflection on the tropic sea were on his cheeks. The blotches ran well down the sides of his face and his hands had the deep-creased scars from handling heavy fish on the cords. But none of these scars were fresh. They were as old as erosions in a fishless desert.	Le vieil homme était maigre et sec, avec des rides comme des coups de couteau sur la nuque. Les taches brunes de cet inoffensif cancer de la peau que cause la réverbération du soleil sur la mer des Tropiques marquaient ses joues ; elles couvraient presque entièrement les deux côtés de son visage ; ses mains portaient les entailles profondes que font les filins au bout desquels se débattent les lourds poissons. Mais aucune de ces entailles n'était récente elles étaient vieilles comme les érosions d'un désert sans poissons.	Le vieil homme était maigre et hâve, avec de profondes rides dans l'arrière du cou. Sur ses joues, les taches brunes d'un cancer de la peau bénin à cause de la réflexion du soleil sur la mer des tropiques. Les taches lui tombaient de chaque côté du visage et ses mains gardaient les cicatrices profondément plissées des poissons hâlés lourdement sur la corde. Mais aucune de ces cicatrices pour être récente. Elles étaient aussi vieilles que l'érosion dans le désert sans poisson.	El viejo era flaco y desgarbado, con arrugas profundas en la parte posterior del cuello. Las pardas manchas del benigno cáncer de la piel que el sol produce con sus reflejos en el mar tropical estaban en sus mejillas. Esas manchas corrían por los lados de su cara hasta bastante abajo y sus manos tenían las hondas cicatrices que causa la manipulación de las cuerdas cuando sujetan los grandes peces. Pero ninguna de estas cicatrices era reciente. Eran tan viejas como las erosiones de un árido desierto.

Figure 275 : Alignement TMX dans un navigateur (avec feuille de styles)



```
<?xml version="1.0" encoding="UTF-8" ?>
<?xml-stylesheet type="text/xsl" href="styles-tmx.xsl"?>
<tmx version="1.4">
 <header adminlang="en" creationdate="20120223T162335Z" creationtool="mkAlign" creationtoolversion="2.00 (2.0b134)" datatype="xml" o-tmf="unknown" segtype="block" srclang="en"/>
<body>
<tu>
<tuv xml:lang="en">
<seg> He was an old man who fished alone in a skiff in the Gulf Stream and he had gone eighty-four days now without taking a fish. In the first forty days a boy had been with him. But after forty days without a fish the boy's parents had told him that the old man was now definitely and finally salao, which is the worst form of unlucky, and the boy had gone at their orders in another boat which caught three good fish the first week. It made the boy sad to see the old man come in each day with his skiff empty and he always went down to help him carry either the coiled lines or the gaff and harpoon and the sail that was furled around the mast. The sail was patched with flour sacks and, furled, it looked like the flag of permanent defeat.
</seg>
</tuv>
<tuv xml:lang="fr1">
<seg> Il était une fois un vieil homme, tout seul dans son bateau qui pêchait au milieu du Gulf Stream. En quatre-vingt-quatre jours, il n'avait pas pris un poisson. Les quarante premiers jours, un jeune garçon l'accompagna ; mais au bout de ce temps, les parents du jeune garçon déclarèrent que le vieux était décidément et sans remède salao ce qui veut dire aussi guignard qu'on peut l'être. On embarqua donc le gamin sur un autre bateau, lequel, en une semaine, ramena trois poissons superbes. Chaque soir le gamin avait la tristesse de voir le vieux rentrer avec sa barque vide. Il ne manquait pas d'aller à sa rencontre et l'aidait à porter les lignes serrées en spirales, la gaffe, le harpon, ou la voile roulée autour du mât. La voile était rapiécée avec de vieux sacs de farine ; ainsi repliée, elle figurait le drapeau en berne de la défaite.
</seg>
</tuv>
<tuv xml:lang="fr2">
<tuv xml:lang="es">
</tu>
<tu>
<tuv xml:lang="en">
<tuv xml:lang="fr1">
<tuv xml:lang="fr2">
<tuv xml:lang="es">
</tu>
```

Figure 276 : Alignement TMX (code source)

La figure précédente donne à voir 2 zones de texte avec pour chacune les 4 volets de langue (dans un ordre identique : en, fr1, fr2, es). Pour la première zone de texte dans le code source, seules les 2 premières langues sont affichées dans l’éditeur.

Un alignement TMX peut être schématisé ainsi :

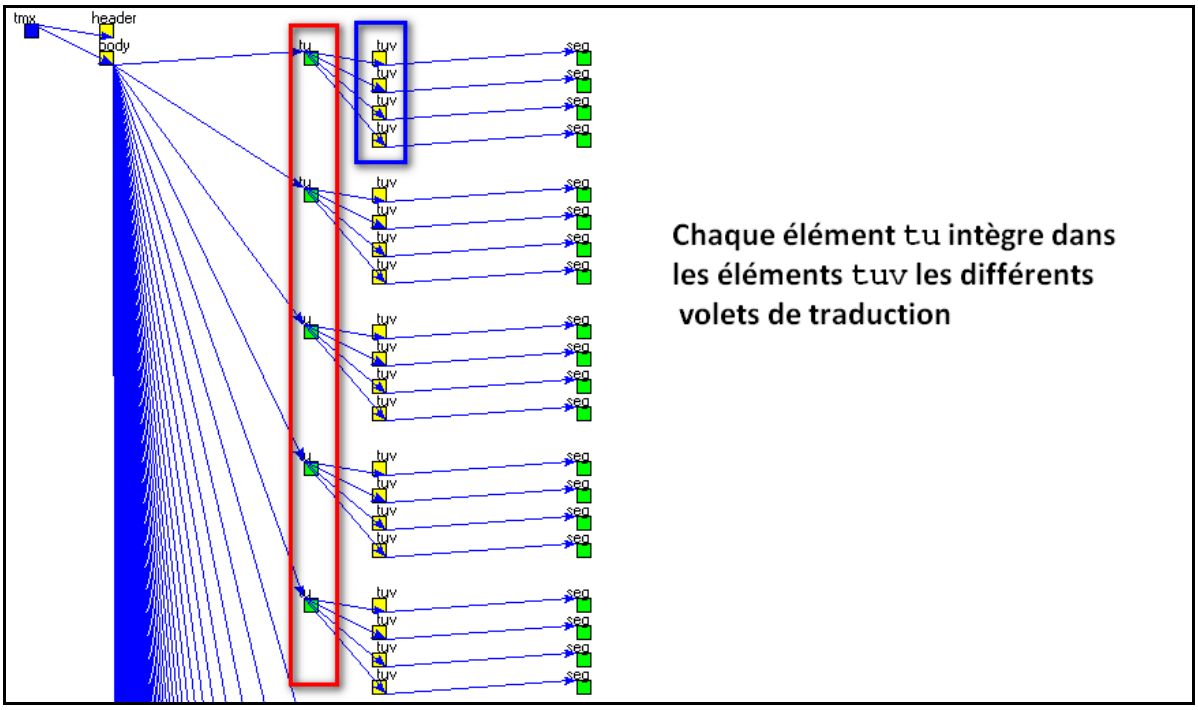


Figure 277 : Alignement TMX , arbre des éléments

Chaque élément `tu` de l'arbre XML encapsule les différents volets de traduction dans autant d'éléments `tuv`. L'alignement se lit de manière verticale pour chaque unité de langue (`tu`) dans la figure précédente.
Pour contraster 2 volets, il faut donc pouvoir mettre au jour les 2 parties visées : pour la source, tous les éléments `tuv` d'une langue donnée contenant la forme visée ; pour la cible, les éléments `tuv` miroirs des précédents dans une autre langue.
Dans le *Trameur*, la *Carte des sections* peut mettre au jour un alignement des différents volets d'un fichier TMX :

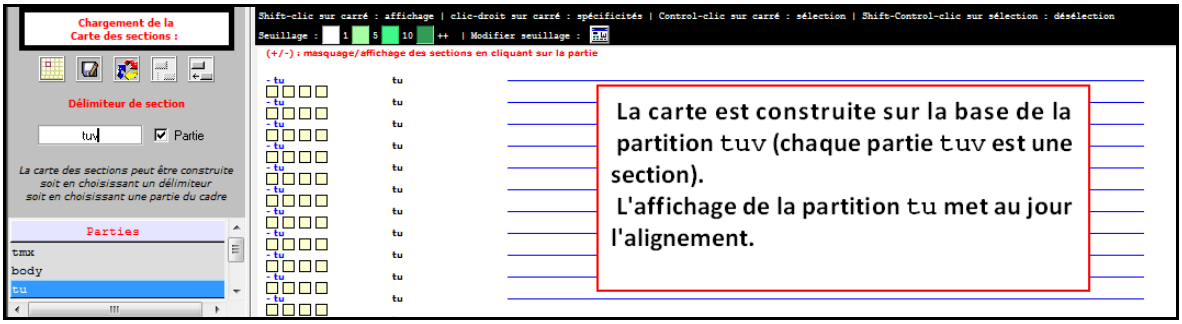

Chargement de la Carte des sections :
Délimiteur de section
tuv
Partie
La carte des sections peut être construite soit en choisissant un délimiteur soit en choisissant une partie du cadre
Parties
tmx
body
tu
La carte est construite sur la base de la partition tuv (chaque partie tuv est une section).
L'affichage de la partition tu met au jour l'alignement.

Figure 278 : Carte des sections d'un alignement TMX avec le Trameur

Ici la *Carte des sections* est générée à partir de la partition `tuv` *i.e* l'ensemble des parties `tuv` de la base textométrique construite à partir de l'alignement TMX. Ces parties intégrées dans le *Cadre* de la base textométrique sont identifiables par leurs positions sur la *Trame* : positions de début et de fin. Dans la carte construite, chaque colonne de sections coïncide avec un volet de l'alignement (1 colonne = 1 langue).

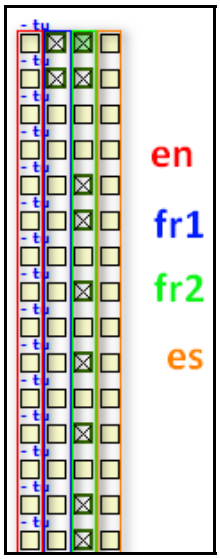


Figure 279 : 1 colonne = 1 volet = 1 langue

Identifier une colonne consiste donc à spécifier le volet visé *i.e* la langue des sections `tuv` de cette colonne. Les langues sont identifiables par les attributs `lang` des balises `tuv` (visibles dans le fichier source *supra*). La *Carte des sections* précédente est construite sur la partition `tuv`, les volets visés doivent être identifiés par la valeur de la langue du volet source et celle du volet cible. Au final, identifier 2 volets distincts consiste à donner un couple de langues parmi celles encodées dans les balises `tuv` du fichier TMX. Une section source donnée ayant une section cible miroir dans un voisinage de sections de longueur égal au nombre de volets de l'alignement (*i.e* le nombre de balises `tuv` dans les éléments `tu`) et les positions respectives d'une section source et d'une section cible restant inchangées tout au long du fichier TMX. Au final, les couples de sections alignées sont aisément identifiables.

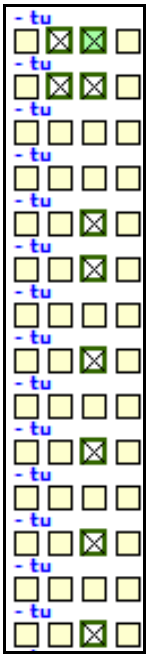


Figure 280 : Distribution d'une forme sur la carte de l'alignement

Une forme étant projetée sur la carte des sections, le calcul de ses cooccurrents consiste à spécifier les 2 volets à contraster. Si on projette la forme “vieil”, a priori, seuls les volets fr2

et fr1 contiennent des sections avec cette forme (colonne 2 et 3 de la carte), la figure précédente donne à voir un extrait de cette distribution.
Le paramétrage du calcul des cooccurrents consiste à définir la valeur de 2 langues à contraster. Nous allons examiner ci-dessous les 2 couples de volets (fr2,fr1) et (fr2,en).

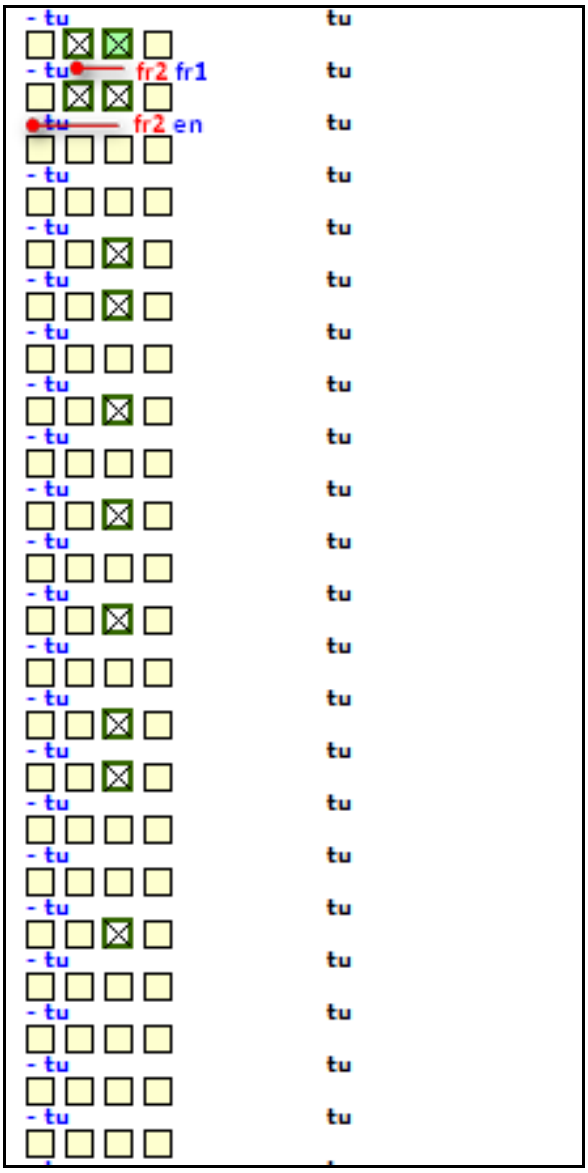


Figure 281 : Contraste de volets (2 à 2)

Pour le couple (fr2, fr1), le paramétrage du calcul est visible ci-dessous : la case à cocher TMX permet de préciser que la base chargée est un alignement au format TMX, les 2 zones de saisie V1 et V2 permettent de spécifier les langues à contraster (attributs `lang` des éléments `tuv`), ici :

```
V1 <= xml:lang=fr2 (SOURCE)
V2 <= xml:lang=fr1(CIBLE)
```

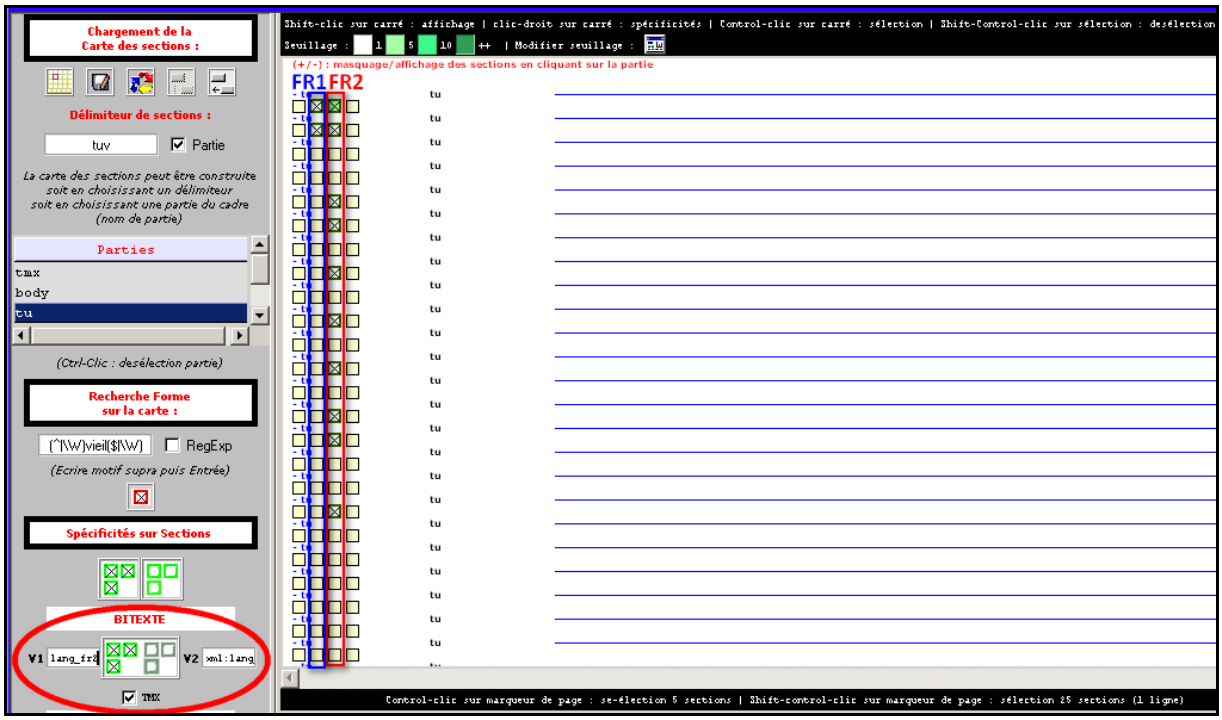

Chargement de la Carte des sections :
Délimiteur de sections :
tuv
Partie
La carte des sections peut être construite soit en choisissant un délimiteur soit en choisissant une partie du cadre (nom de partie)
Parties
tmx
body
tu
(Ctrl-Clic : desélection partie)
Recherche Forme sur la carte :
(^|\W)vieil($|\W)
RegExp
(Ecrire motif supra puis Entrée)
Spécificités sur Sections
BITEXTE
¥1 lang_fr
¥2 xml:lang
TMX
Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : desélection
Seuillage :
Modifier seuillage :
(+/-) : masquage/affichage des sections en cliquant sur la partie
FR1 FR2
Control-clic sur marqueur de page : se-élection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)

Figure 282 : Contraste de 2 volets (FR2, FR1)

Et le résultat du calcul a l'allure suivante :

Specificites du vocabulaire (Annotation Forme) sur les sections SOURCE contenant le motif : vieil

```
Nombre d'occurrences du texte global = 29257
Nombre d'occurrences dans la partie visee = 10191
Seuil : 10
---------------------------------------- SOURCE ----------------------------------------------------
---------------------------------------- (cible infra) ---------------------------------------------
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit :
C:/DOCUME~1/FLEURY~1/MESDOC~1/MYDROP~1/WIP/WIP-TR~1/ALIGNE~1/ALIGNE~1/export/resultspecif-source-1415131422.txt
```

Forme	Ind-Specif	Fq-Totale	Fq-Partie
vieil	**	185	185
homme	**	211	189
Le	11.6	125	80
le	8.8	1005	434
requin	6.9	40	29
sur	4.7	175	84
ses	4.6	106	55
lame	4.3	10	9
était	4.3	251	113
étaient	4.0	40	24
yeux	3.9	28	18
nageoires	3.9	9	8
avaient	3.9	32	20
avait	3.8	177	81
parlaient	3.7	6	6
frappa	3.6	13	10
dents	3.5	8	7
continuait	3.3	10	8
Parfois	3.3	5	5
harpon	3.2	20	13
ainsi	3.1	7	6
surface	3.1	29	17
et	3.0	989	379
leurs	3.0	19	12
son	3.0	157	69
poissons	3.0	34	19
eau	2.9	73	35
chair	2.9	15	10
salée	2.8	4	4
élevait	2.8	4	4

Figure 283 : Spécificités volet SOURCE (FR2)

---------- CIBLE ----------

Specificites du vocabulaire (Annotation Forme) sur les sections CIBLE miroir de celles contenant le motif : vieil

Nombre d'occurrences du texte global = 29982
Nombre d'occurrences dans la partie visee = 10133
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)

Forme	Ind-Specif	Fq-Totale	Fq-Partie
vieux	32.8	269	186
Le	11.3	181	104
requin	7.7	35	27
vieil	7.0	16	15
le	5.6	889	358
avait	4.9	163	78
sa	4.7	123	61
la	4.5	826	326
du	4.1	209	93
lame	3.8	6	6
l	3.7	525	209
surface	3.7	15	11
eau	3.7	80	40
harpon	3.7	19	13
ses	3.6	98	47
étaient	3.6	43	24
Les	3.5	46	25
frappa	3.5	8	7
jaillissaient	3.4	5	5
chair	3.4	12	9
grosses	3.4	5	5
était	3.2	251	104
plusieurs	3.1	7	6
gueule	3.0	9	7
mâchoires	3.0	9	7
aperçut	3.0	17	11
côté	3.0	17	11
vit	3.0	11	8
laissait	2.9	4	4
faux	2.9	4	4

Figure 284 : Spécificités volet CIBLE (FR1)

On retrouve ci-dessus les résultats produits précédemment.
De même, pour le couple (fr2, en), le paramétrage du calcul est visible ci-dessous :

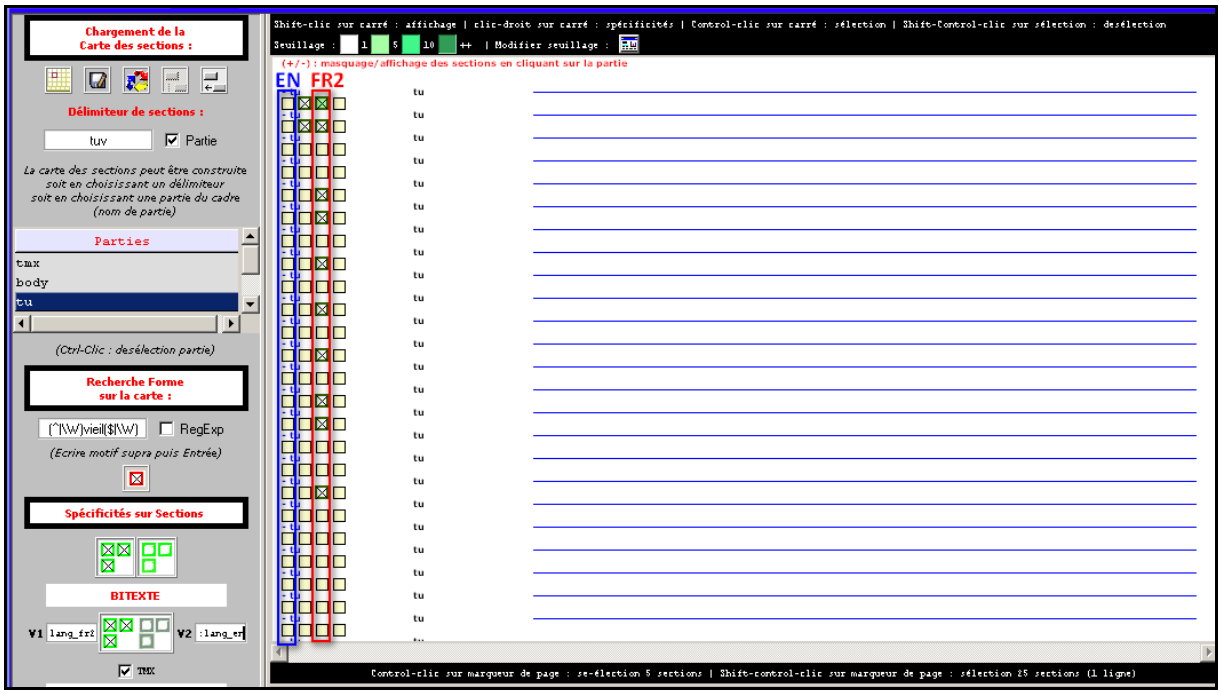

Chargement de la Carte des sections :
Délimiteur de sections :
tuv
Partie
La carte des sections peut être construite soit en choisissant un délimiteur soit en choisissant une partie du cadre (nom de partie)
Parties
tmx
body
tu
(Ctrl-Clic : desélection partie)
Recherche Forme sur la carte :
(?i)\W)vieil($|\W)
RegExp
(Ecrire motif supra puis Entrée)
Spécificités sur Sections
BITEXTE
V1 lang_fr2
V2 :lang_en
tmx
Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : desélection
Seuillage : 1 5 10 ++ | Modifier seuillage :
(+/-) : masquage/affichage des sections en cliquant sur la partie
EN FR2
tu
Control-clic sur marqueur de page : sé-élection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)

Figure 285 : Contraste de 2 volets (FR2, EN)

Et le résultat produit a l'allure suivante :

Specificites du vocabulaire (Annotation Forme) sur les sections SOURCE contenant le motif : vieil

Nombre d'occurrences du texte global = 29257
Nombre d'occurrences dans la partie visee = 10191
Seuil : 10
---------- SOURCE ----------
---------- (cible infra) ----------
(Specificites positives en haut de liste, negatives en bas)
Le fichier construit :
C:/DOCUME~1/FLEURY~1/MESDOC~1/MYDROP~1/WIP/WIP-TR~1/ALIGNE~1/ALIGNE~1/export/resultspecif-source-1415131664.txt

Forme	Ind-Specif	Fq-Totale	Fq-Partie
vieil	**	185	185
homme	**	211	189
Le	11.6	125	80
le	8.8	1005	434
requin	6.9	40	29
sur	4.7	175	84
ses	4.6	106	55
lame	4.3	10	9
était	4.3	251	113
étaient	4.0	40	24
yeux	3.9	28	18
nageoires	3.9	9	8
avaient	3.9	32	20
avait	3.8	177	81
parlaient	3.7	6	6
frappa	3.6	13	10
dents	3.5	8	7
continuait	3.3	10	8
Parfois	3.3	5	5
harpon	3.2	20	13
ainsi	3.1	7	6
surface	3.1	29	17
et	3.0	989	379
leurs	3.0	19	12
son	3.0	157	69
poissons	3.0	34	19
eau	2.9	73	35

Figure 286 : Spécificités volet SOURCE (FR2)

---------- CIBLE ----------
Specificites du vocabulaire (Annotation Forme) sur les sections CIBLE miroir de celles contenant le motif : vieil

Nombre d'occurrences du texte global = 26847
Nombre d'occurrences dans la partie visee = 9512
Seuil : 10
(Specificites positives en haut de liste, negatives en bas)

Forme	Ind-Specif	Fq-Totale	Fq-Partie
old	49.4	247	200
man	40.6	264	199
The	12.1	194	115
the	7.6	2119	858
shark	7.2	46	33
his	6.3	430	197
her	6.0	29	22
water	4.7	106	56
were	4.5	138	69
wide	4.4	13	11
was	4.4	435	188
hit	4.3	26	18
blade	4.2	10	9
had	3.8	203	93
came	3.7	41	24
as	3.6	191	87
by	3.5	29	18
jaws	3.5	17	12
turtles	3.4	8	7
s	3.4	77	39
neck	3.3	5	5
afraid	3.3	5	5
flesh	3.3	5	5
side	3.3	32	19
rose	3.2	14	10
harpoon	3.2	20	13
surface	3.2	16	11
and	3.0	1233	475
closed	3.0	7	6
birds	3.0	7	6
they	3.0	99	47
Sometimes	3.0	7	6
coming	2.9	15	10

Figure 287 : Spécificités volet CIBLE (EN)

30.2.2.4 Importer une base « d'alignements » multi-annotée

Les calculs présentés ci-dessus peuvent être réalisés sur tous les niveaux d'annotation disponibles dans la base textométrique chargée dans le ***Trameur***.

Nous allons illustrer cette fonctionnalité en construisant une base multi-annotée et intégrant 2 volets alignés. Pour cela, nous procédons de la manière suivante :

- Chargement avec le ***Trameur*** de la base complète précédente (convention FR et EN) avec étiquetage en Français via *treetagger*
- Export de la base précédente : création d'une base multi-annotée (BASE-FR-EN) d'un bitexte, le volet FR est « bien » annoté, le volet EN ne l'est pas.
- Chargement avec le ***Trameur*** d'une base constituée uniquement du volet EN avec étiquetage en Anglais via *treetagger*
- Export de la base précédente : création d'une base multi-annotée du volet EN (BASE-EN)
- Dans un éditeur, suppression dans la base BASE-FR-EN de la *Trame* correspondant au volet EN et insertion « au même endroit » de la *Trame* de la base BASE-EN. Dans la base BASE-FR-EN, les 2 volets sont désormais « bien » annotés (chaque volet ayant son propre jeu d'annotations)

La base BASE-FR-EN peut finalement être importée dans le ***Trameur***

La base résultante est visible dans la figure suivante (2 extraits de la *Trame* dans lesquels on donne à voir le début des 2 volets) : chaque item de la *Trame* dispose de 3 niveaux d'annotation : forme (balise <f>), lemme (balise <l>) et POS (balise (<c>).

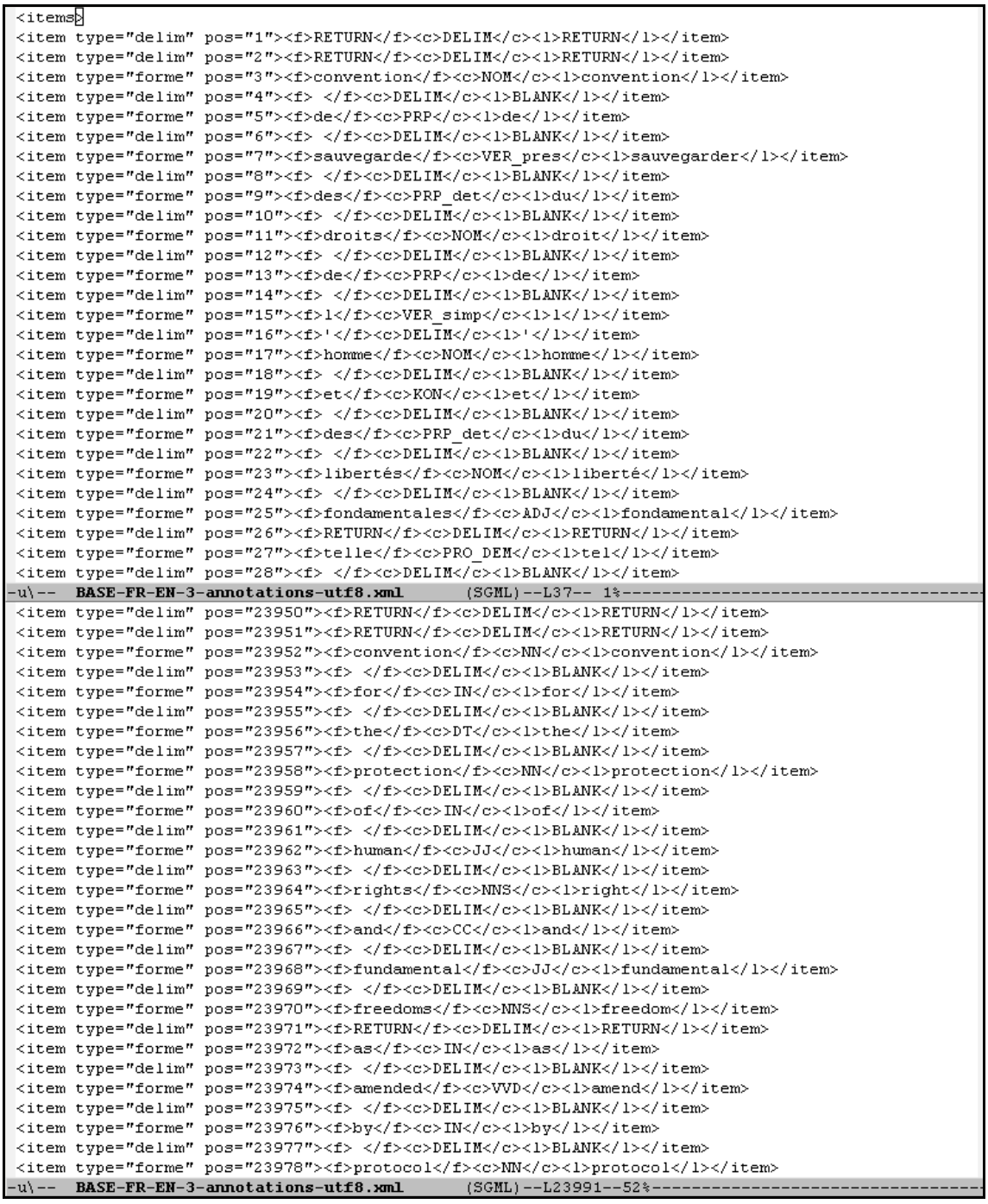


Figure 288 : Base textométrique multiannotée d'un alignement de textes bilingues

Une fois importée, les mêmes calculs peuvent être réalisés sur la *Carte des sections* mettant au jour l'alignement.

On pourrait bien entendu disposer d'une base multi-annotée et intégrant 2 volets alignés en procédant d'une autre manière : on pourrait par exemple projeter (via un programme ou à la main) des annotations sur la *Trame* d'une base intégrant 2 volets alignés.

30.2.2.5 Calculs contrastifs sur une base textométrique de textes alignés

Au delà du calcul des cooccurrents sur 2 volets alignés d'une base textométrique *via* la *Carte des sections*, il est possible de mener des calculs textométriques sur chacun des volets en utilisant le *Gestionnaire de Sélection* et notamment les modules permettant de lancer des calculs sur telle ou telle partie de la base de travail.

Sur la base précédente par exemple, les modules de calcul encadrés dans la figure suivante peuvent être lancés sur un des volets de la base après sélection du volet visé dans la liste des parties de la partition volet.

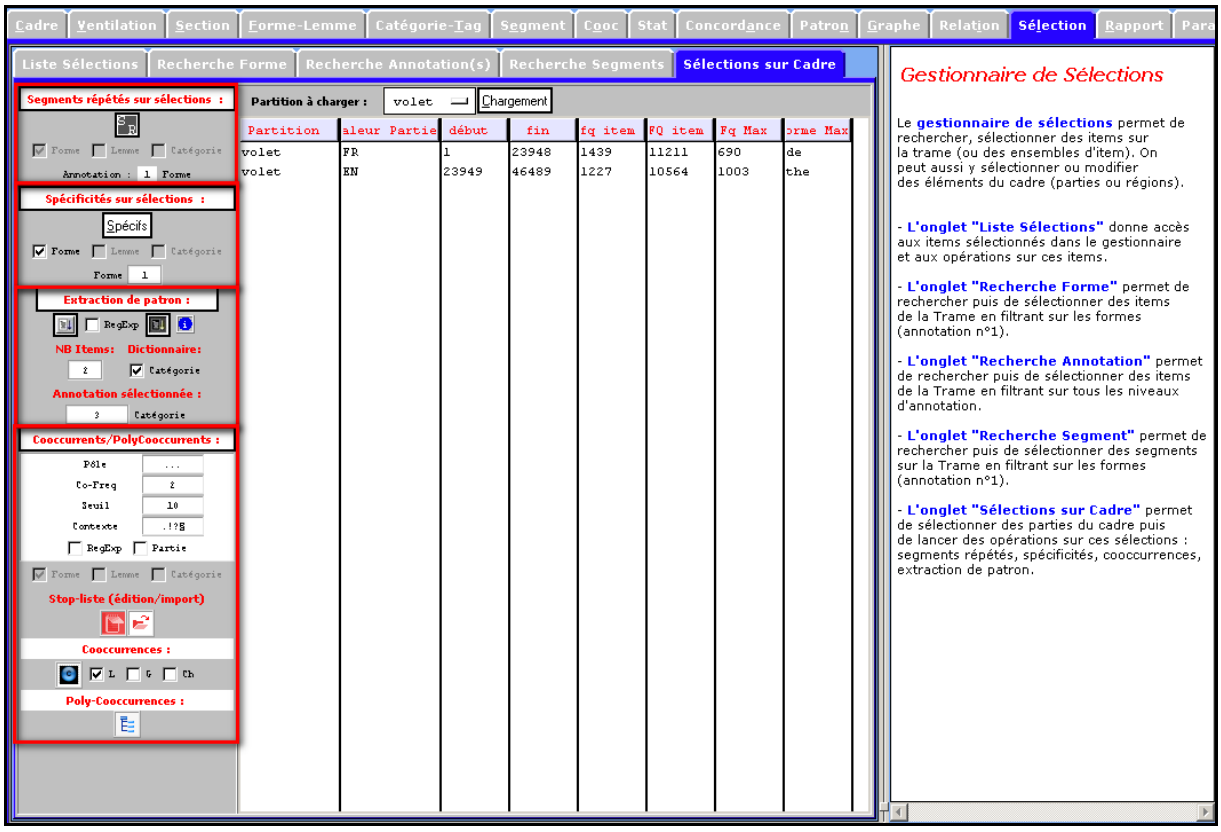


Figure 289 : Calculs contrastifs sur une base textométrique d'alignements

30.2.3 Outils et méthodes d'exploration d'un alignement

Dans la suite de ce document, 2 bases textométriques sont utilisées :

1. **Base bilingue ParTUT2Trameur** (français/anglais). Les 2 volets de cette base disposent de 9 niveaux d'annotation (Alignement de treebank). Cette base permet de tester les traitements spécifiques pour les bases textométriques intégrant 2 textes. **Alignement de Treebank pour le Trameur** (pdf) : présentation du processus de transcodage de 2 treebanks ParTUT pour construire une base Textométrique.
2. **Base bilingue "Investiture Obama"** (français/anglais). Cette base regroupe le discours original en anglais prononcé par B. Obama le 20 janvier 2009 à Washington, publié sur le site de The New York Times (volet EN), et 4 traductions françaises de ce discours (volets FR0-1-2-3). Les traductions ont été récupérées sur le site officiel de la Maison Blanche (volet FR0), sur les sites des journaux français Le Monde (volet FR1) et Libération (volet FR2), ainsi que sur le site de RFI (volet FR3). Les 5 volets de cette base disposent de 3 niveaux d'annotation (forme, lemme, catégorie). Cette base permet de tester les traitements spécifiques pour les bases textométriques intégrant 2 textes alignés.

30.2.3.1 Edition d'un alignement multi-volets

La base *Investiture* étant chargée (par importation), la carte des sections construite via le délimiteur de section § a l'allure suivante :

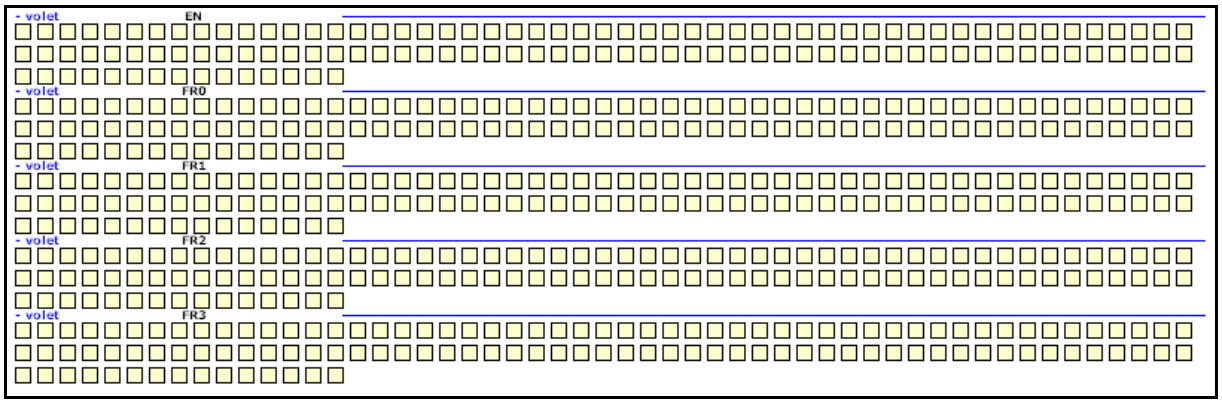


Figure 290 : Carte de sections d'un alignement multi-volets

L'alignement des 5 volets se lit de manière verticale comme dans les exemples précédents. Sous la carte des sections, une case à cocher et une zone de saisie permettent de modifier l'affichage des sections. Par défaut, un clic sur une section affiche le contenu de la section dans l'éditeur de section. Si la case à cocher Bitext est cochée et si on donne le nombre de volets présents dans l'alignement chargé et mis au jour dans la carte des sections :

Dans ce cas, l'édition de la section cliquée se fait en insérant aussi ses sections miroirs.

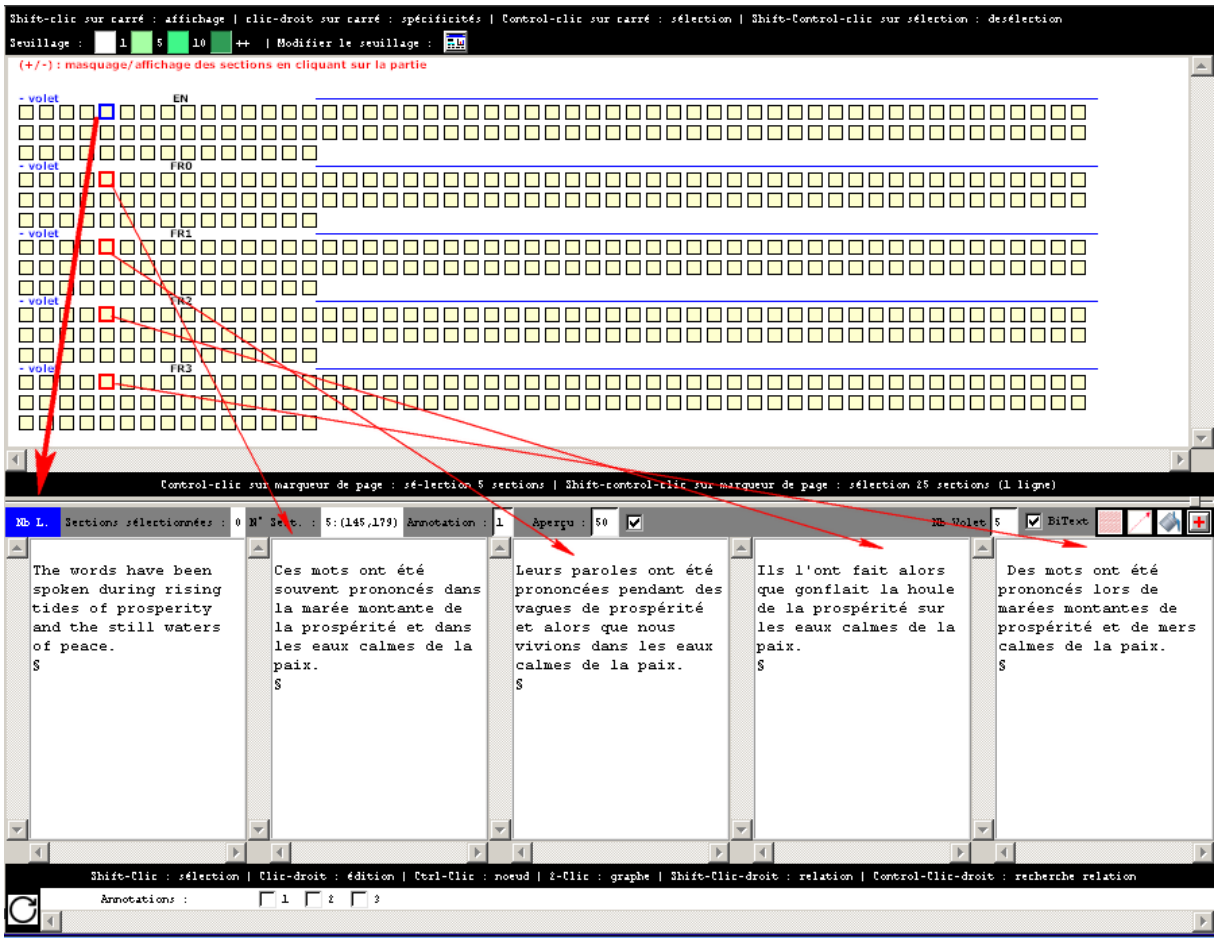

Shift-clic sur carré : affichage | clic-droit sur carré : spécificités | Control-clic sur carré : sélection | Shift-Control-clic sur sélection : desélection
Seuillage :
(+/-) : masquage/affichage des sections en cliquant sur la partie
- volet EN
- volet FR0
- volet FR1
- volet FR2
- volet FR3
Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)
Nb L.
Sections sélectionnées : 0
Annotation : 1
Aperçu : 50
Nb Volet 5
BiText
The words have been spoken during rising tides of prosperity and the still waters of peace.
$
Ces mots ont été souvent prononcés dans la marée montante de la prospérité et dans les eaux calmes de la paix.
$
Leurs paroles ont été prononcées pendant des vagues de prospérité et alors que nous vivions dans les eaux calmes de la paix.
$
Ils l'ont fait alors que gonflait la houle de la prospérité sur les eaux calmes de la paix.
$
Des mots ont été prononcés lors de marées montantes de prospérité et de mers calmes de la paix.
$
Shift-Clic : sélection | Clic-droit : édition | Ctrl-Clic : noeud | 2-Clic : graphe | Shift-Clic-droit : relation | Control-Clic-droit : recherche relation
Annotations :

Figure 291 : Edition d'un alignement multi-volets

Dans la figure précédente, le clic sur la section du volet EN (carré bleu) déclenche la mise au jour des sections miroirs et l'affichage en parallèle de tous les contenus de ces sections alignées.

Remarque : la section cliquée dans la carte est éditée sur la gauche de l'éditeur multi-volets, ses miroirs sont ensuite éditées sur sa droite dans l'ordre d'apparition dans la carte.

30.3.3.2 Opérations dans une section multi-volets

Toutes les opérations disponibles dans l'éditeur de section peuvent désormais être réalisées sur une section multi-volets. Ci-dessous, l'affichage de 3 niveaux d'annotations sur les 5 sections éditées conjointement :

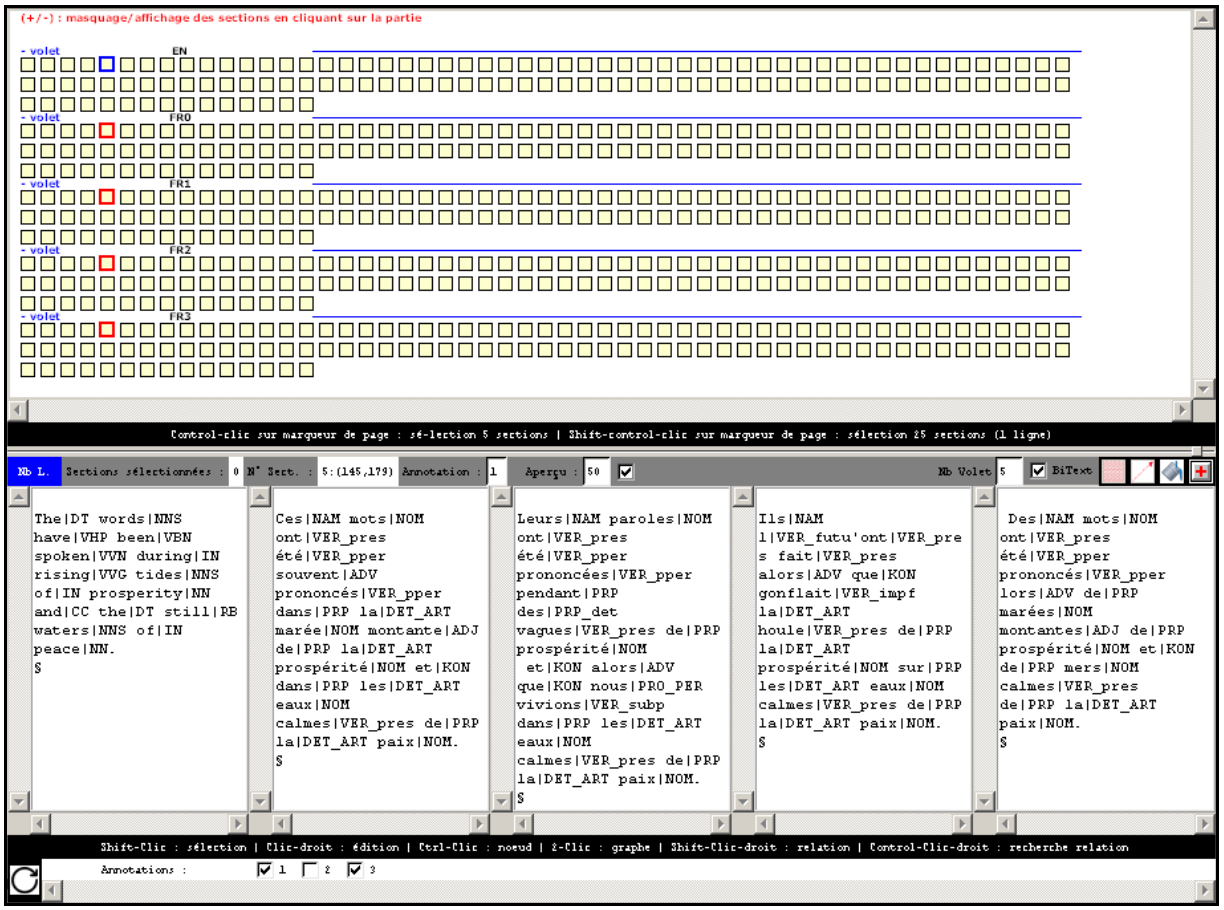

(+/-) : masquage/affichage des sections en cliquant sur la partie
- volet EN
- volet FR0
- volet FR1
- volet FR2
- volet FR3
Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)
Nb L.
Sections sélectionnées : 0
N° Sect. : 5:(145,179)
Annotation : 1
Aperçu : 50
Nb Volet 5
BiText
The|DT words|NNS have|VHP been|VBN spoken|VVN during|IN rising|VVG tides|NNS of|IN prosperity|NN and|CC the|DT still|RB waters|NNS of|IN peace|NN. $
Ces|NAM mots|NOM ont|VER_pres été|VER_pper souvent|ADV prononcés|VER_pper dans|PRP la|DET_ART marée|NOM montante|ADJ de|PRP la|DET_ART prospérité|NOM et|KON dans|PRP les|DET_ART eaux|NOM calmes|VER_pres de|PRP la|DET_ART paix|NOM. $
Leurs|NAM paroles|NOM ont|VER_pres été|VER_pper prononcées|VER_pper pendant|PRP des|PRP_det vagues|VER_pres de|PRP prospérité|NOM et|KON alors|ADV que|KON nous|PRO_PER vivions|VER_subp dans|PRP les|DET_ART eaux|NOM calmes|VER_pres de|PRP la|DET_ART paix|NOM. $
Ils|NAM l|VER_futu'ont|VER_pres fait|VER_pres alors|ADV que|KON gonflait|VER_impf la|DET_ART houle|VER_pres de|PRP la|DET_ART prospérité|NOM sur|PRP les|DET_ART eaux|NOM calmes|VER_pres de|PRP la|DET_ART paix|NOM. $
Des|NAM mots|NOM ont|VER_pres été|VER_pper prononcés|VER_pper lors|ADV de|PRP marées|NOM montantes|ADJ de|PRP prospérité|NOM et|KON de|PRP mers|NOM calmes|VER_pres de|PRP la|DET_ART paix|NOM. $
Shift-Clic : sélection | Clic-droit : édition | Ctrl-Clic : noeud | 2-Clic : graphe | Shift-Clic-droit : relation | Control-Clic-droit : recherche relation
Annotations : 1 2 3

Figure 292 : Opérations dans une section multi-volets

30.3.3.3 Graphe des relations de dépendances sur une section multi-volets

Cette fois-ci, on considère que la base *parTUT* est chargée dans le Trameur. Cette base intègre un alignement de *treebanks* (*cf* présentation en ligne de cette ressource) : un volet EN et un volet FR. On édite conjointement une section et son miroir :

- PARTIE EN
- PARTIE FR
Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)
Nb L.
Sections sélectionnées : 0
N° Sect. : 106:(5133,5157)
Annotation : 1
Aperçu : 50
Nb Volet 2
BiText
(The order of business was adopted thus amended) . $
(Le Parlement adopte l' ordre des des travaux ainsi modifié) . $

Figure 293 : Graphe de relations de dépendances dans une section multi-volets (1)

Le bouton « Affichage des relations de dépendance dans la section » (cerclé de rouge ci-dessus) permet de mettre au jour le graphe de relations sur les 2 zones textuelles alignées ici :

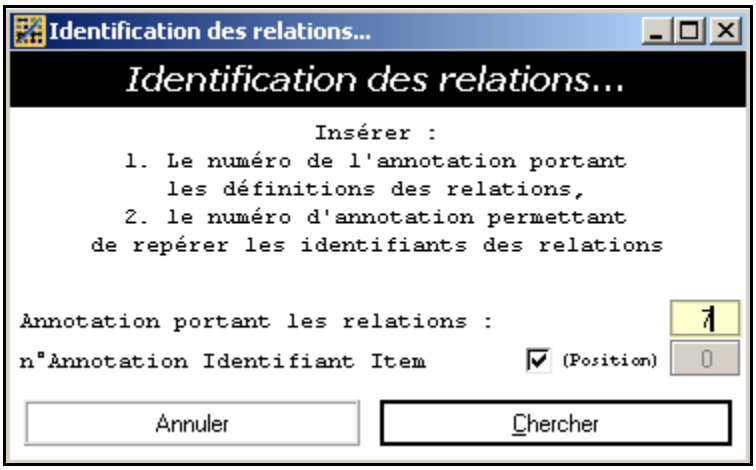


Figure 294 : Graphe de relations de dépendances dans une section multi-volets (2)

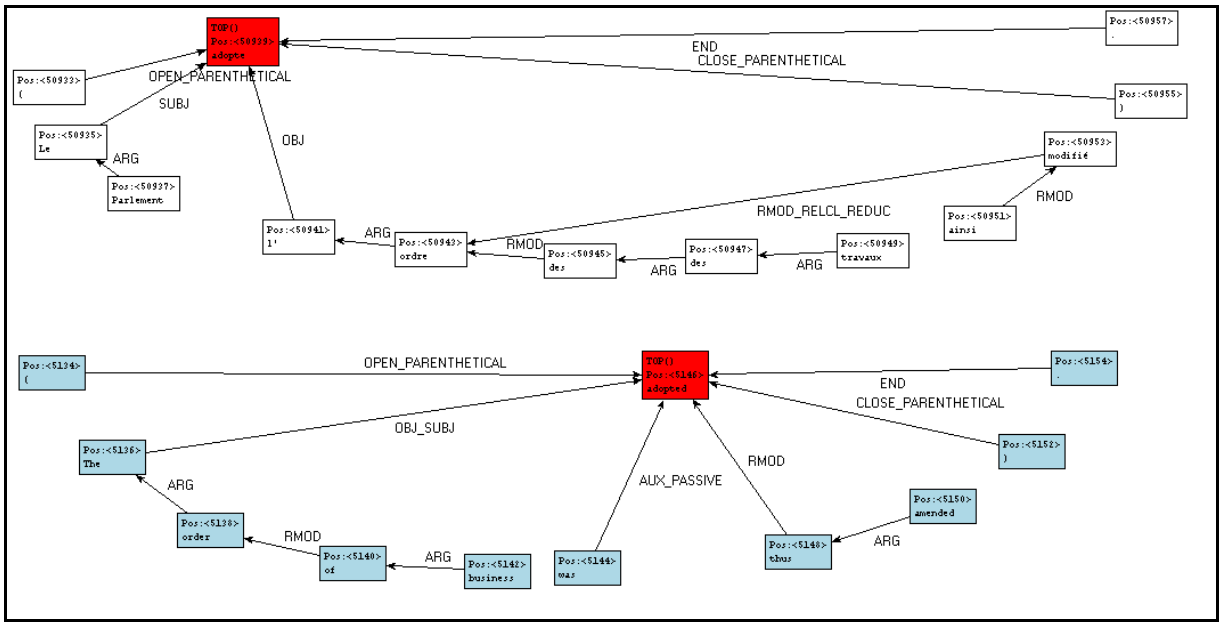


Figure 295 : Graphe de relations de dépendances dans une section multi-volets (3)

30.3 Annotation de données orales

Projet associé : CFPP2000 http://cfpp2000.univ-paris3.fr/index.html

Un moteur de recherche[27] permet de lancer des requêtes dans les transcriptions de **CFPP2000** annotées via *treetagger* (similaire à celui déjà en place sur le texte brut des transcriptions[28]). Les données ont été annotées en utilisant les ressources développées pour le traitement de l'oral via *treetagger* par Christophe Benzitoun (ATILF) et intégrées dans le ***Trameur***.

Après chargement et annotation d'un fichier dans le ***Trameur***, on peut exporter la "base annotée" résultante dans un format XML « proche » du format de départ (accessible via le bouton) : la base résultante est de fait le fichier de départ sur lequel on a projeté des annotations construites par *treetagger*.

Pour le moment les balises auto-fermantes ne sont pas gérées dans le ***Trameur***, elles ne sont dans pas conservées dans cet export. Ce type de balise est utilisé dans les textes de transcriptions construits par *Transcriber* par exemple (outil utilisé pour la transcription dans le corpus **CFPP2000**) pour noter des commentaires, distinguer les locuteurs quand plusieurs sont associés à un seul tour de paroles etc. Les figures suivantes permettent de visualiser la perte d'information induite : si on cherche la forme "fonctionnaire" dans les transcriptions brutes puis dans les transcriptions annotées on obtient les 2 résultats suivants :

Recherche sur texte brut :

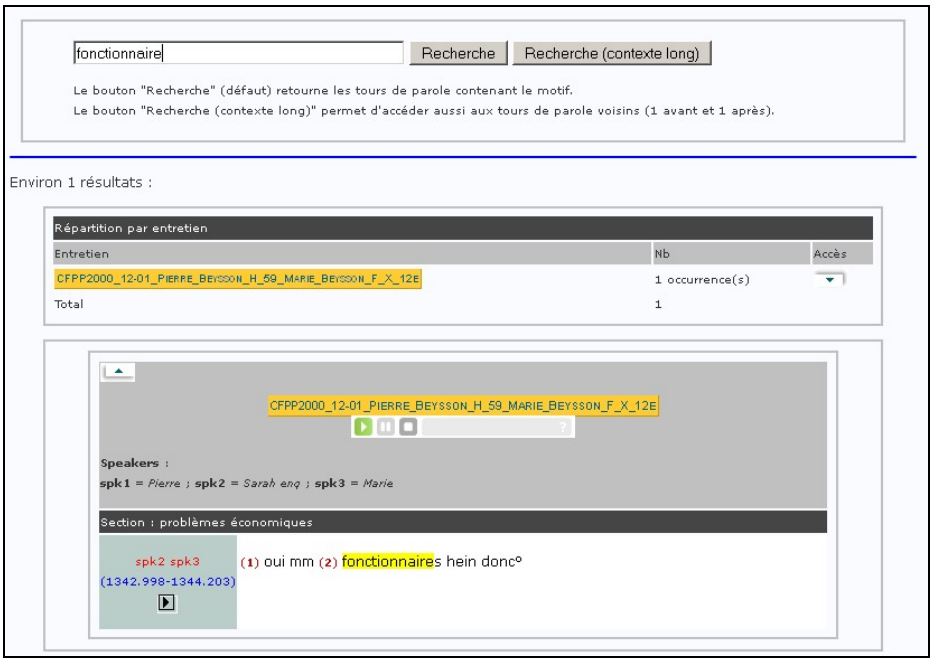


Figure 296 : CFPP2000 (1)

Recherche sur texte annoté : (on perd la distinction spk2/spk3)

[27] http://cfpp2000.univ-paris3.fr/search-transcription-tt/

[28] http://cfpp2000.univ-paris3.fr/search-transcription/

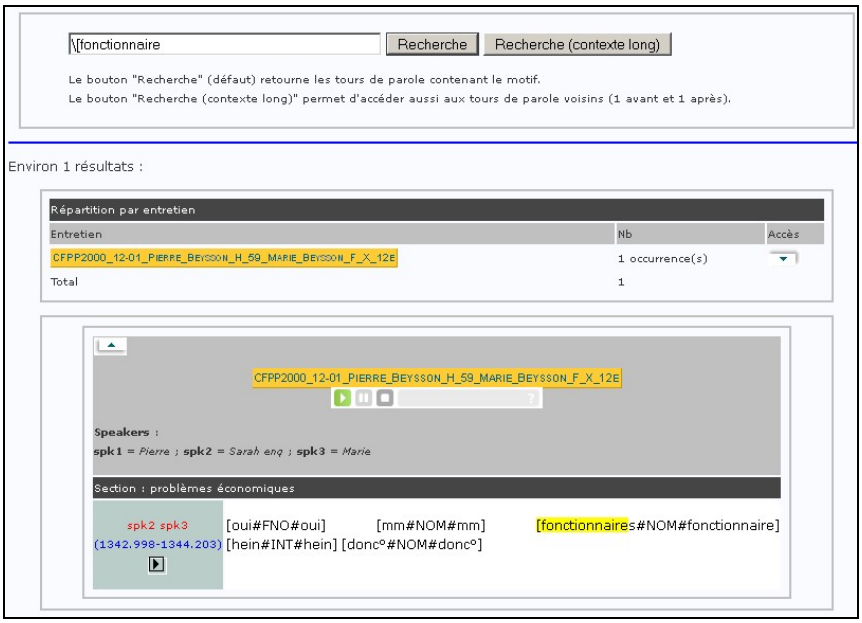


Figure 297 : CFPP2000 (2)

Dans le cas de **CFPP2000**, le processus (d'annotation puis d'export) intégré dans le ***Trameur*** permet d'annoter facilement des fichiers de transcription issus de *Transcriber* par exemple et de récupérer ensuite les mêmes fichiers avec annotations morphosyntaxiques (pour le moment, il reste à faire quelques modifications mineures sur le fichier exporté). On peut donc très facilement et à moindre coût réitérer l'étiquetage et produire un nouveau fichier annoté que l'on peut ensuite intégrer dans le moteur de recherche dans les données annotées de **CFPP2000**.

Les éventuelles corrections sur les annotations produites par *treetagger* peuvent se faire dans le ***Trameur*** (*cf* processus de correction des annotations présentés dans ce document) ou de manière externe (*via* un éditeur XML par exemple).

30.4 Module petiMoteur

Le module *petiMoteur* est disponible dans l'onglet **CADRE**. Ce module permet de réaliser un corpus à partir d'une liste d'URLs (ici une liste de 60 urls pointant vers des pages d'accueil de journaux français en ligne). Une fois le fichier d'URLs chargé (1), le module d'aspiration (2) construit un fichier concaténant les contenus textuels des différentes URLs. Le fichier produit (text_URLS.txt dans le même dossier que le fichier des URLs) peut ensuite être chargé dans le ***Trameur***.

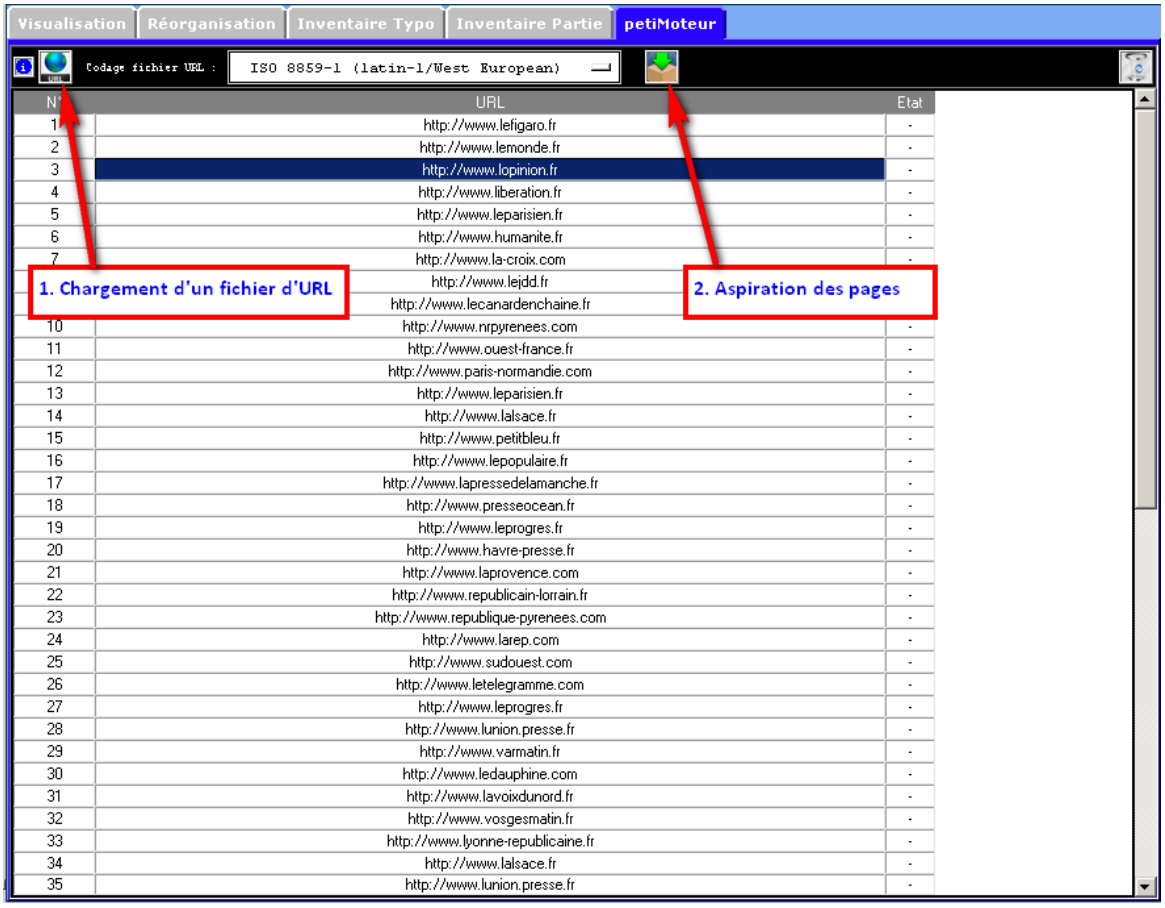


Figure 298 : Le petiMoteur

31 Le Trameur en mode console

Le ***Trameur*** en mode console s'utilise dans une fenêtre de commandes :

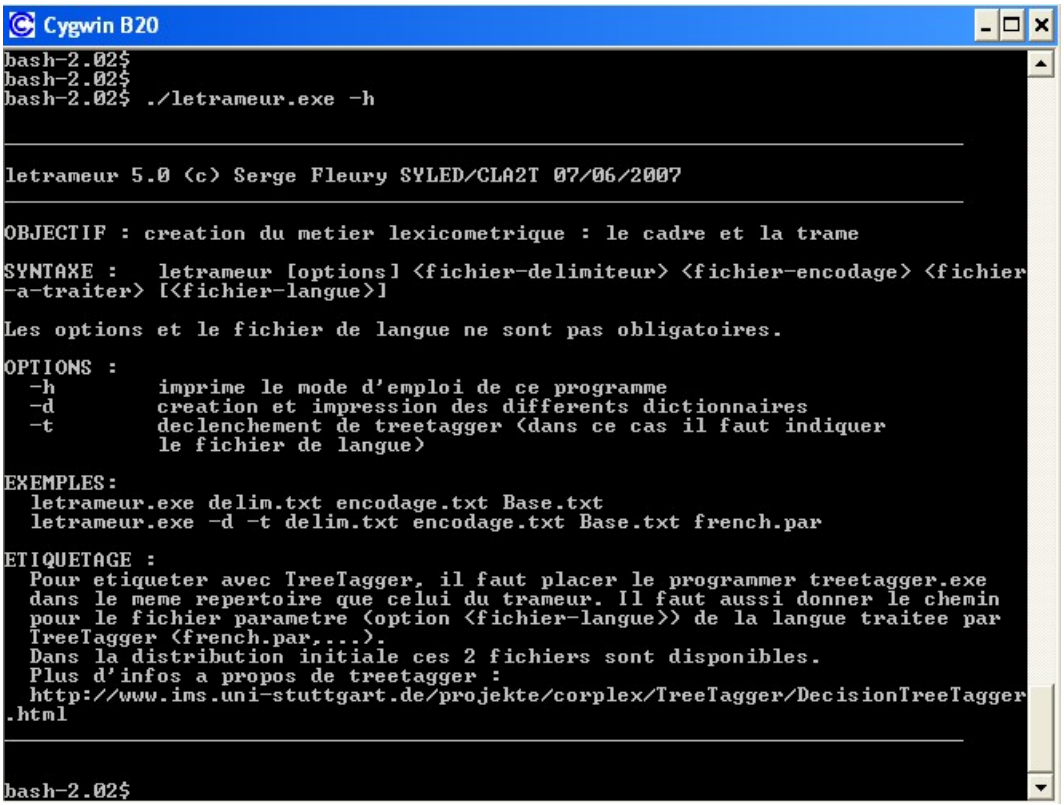

Cygwin B20
bash-2.02$
bash-2.02$
bash-2.02$ ./letrameur.exe -h
letrameur 5.0 (c) Serge Fleury SYLED/CLA2T 07/06/2007
OBJECTIF : creation du metier lexicometrique : le cadre et la trame
SYNTAXE : letrameur [options] <fichier-delimiteur> <fichier-encodage> <fichier-a-traiter> [<fichier-langue>]
Les options et le fichier de langue ne sont pas obligatoires.
OPTIONS :
-h imprime le mode d'emploi de ce programme
-d creation et impression des differents dictionnaires
-t declenchement de treetagger (dans ce cas il faut indiquer le fichier de langue)
EXEMPLES:
letrameur.exe delim.txt encodage.txt Base.txt
letrameur.exe -d -t delim.txt encodage.txt Base.txt french.par
ETIQUETAGE :
Pour etiqueter avec TreeTagger, il faut placer le programmer treetagger.exe dans le meme repertoire que celui du trameur. Il faut aussi donner le chemin pour le fichier parametre (option <fichier-langue>) de la langue traitee par TreeTagger (french.par,...).
Dans la distribution initiale ces 2 fichiers sont disponibles.
Plus d'infos a propos de treetagger :
http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html
bash-2.02$

Figure 299 : *Le Trameur* en mode console

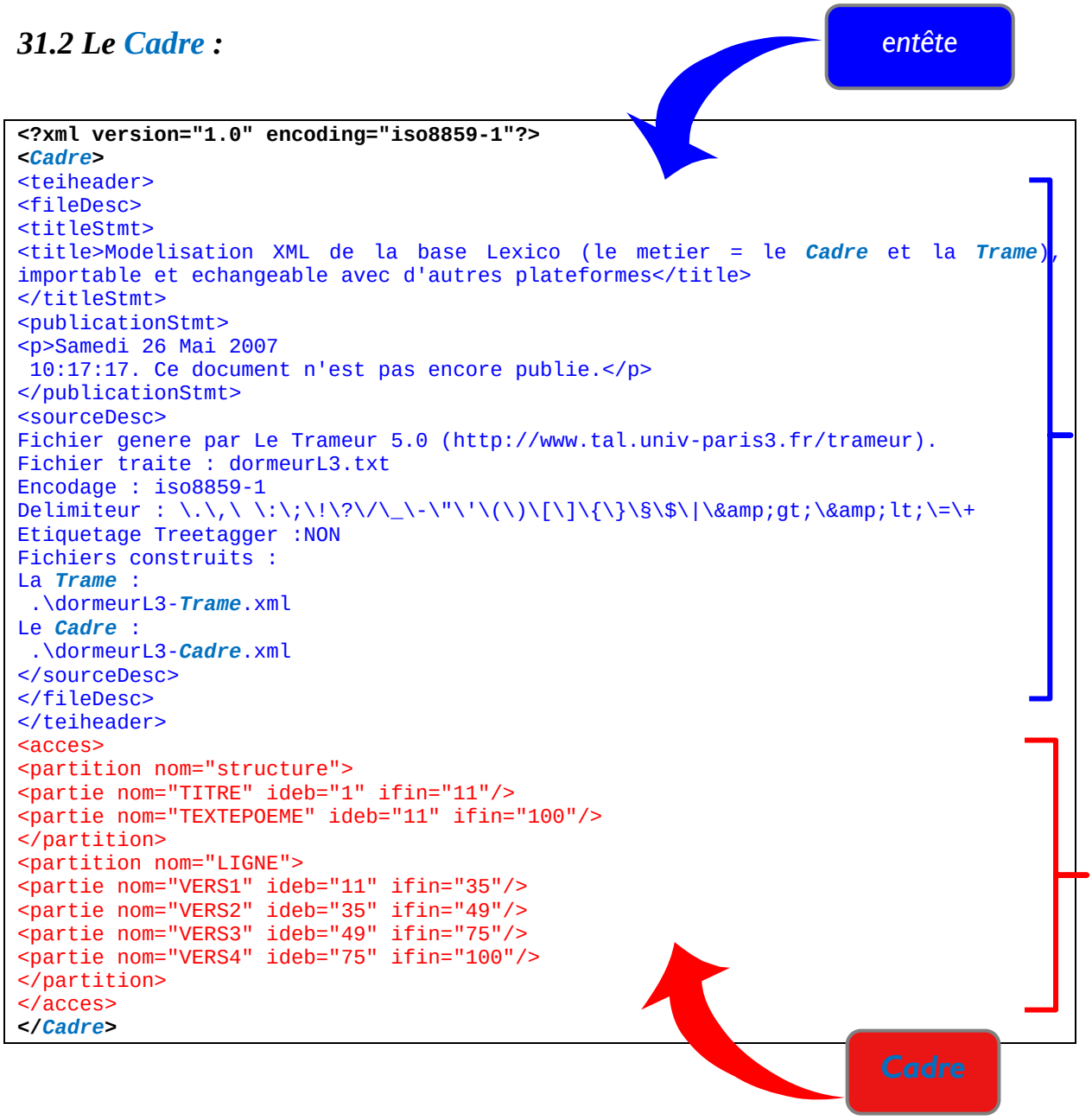
Le seul objectif de ce programme est de construire le ***Cadre*** et la ***Trame*** d'un fichier.

Nous allons montrer sur le même fichier de travail les différents fichiers produits suivant les options définies et visibles dans la figure précédente.

31.1 Fichier de travail :

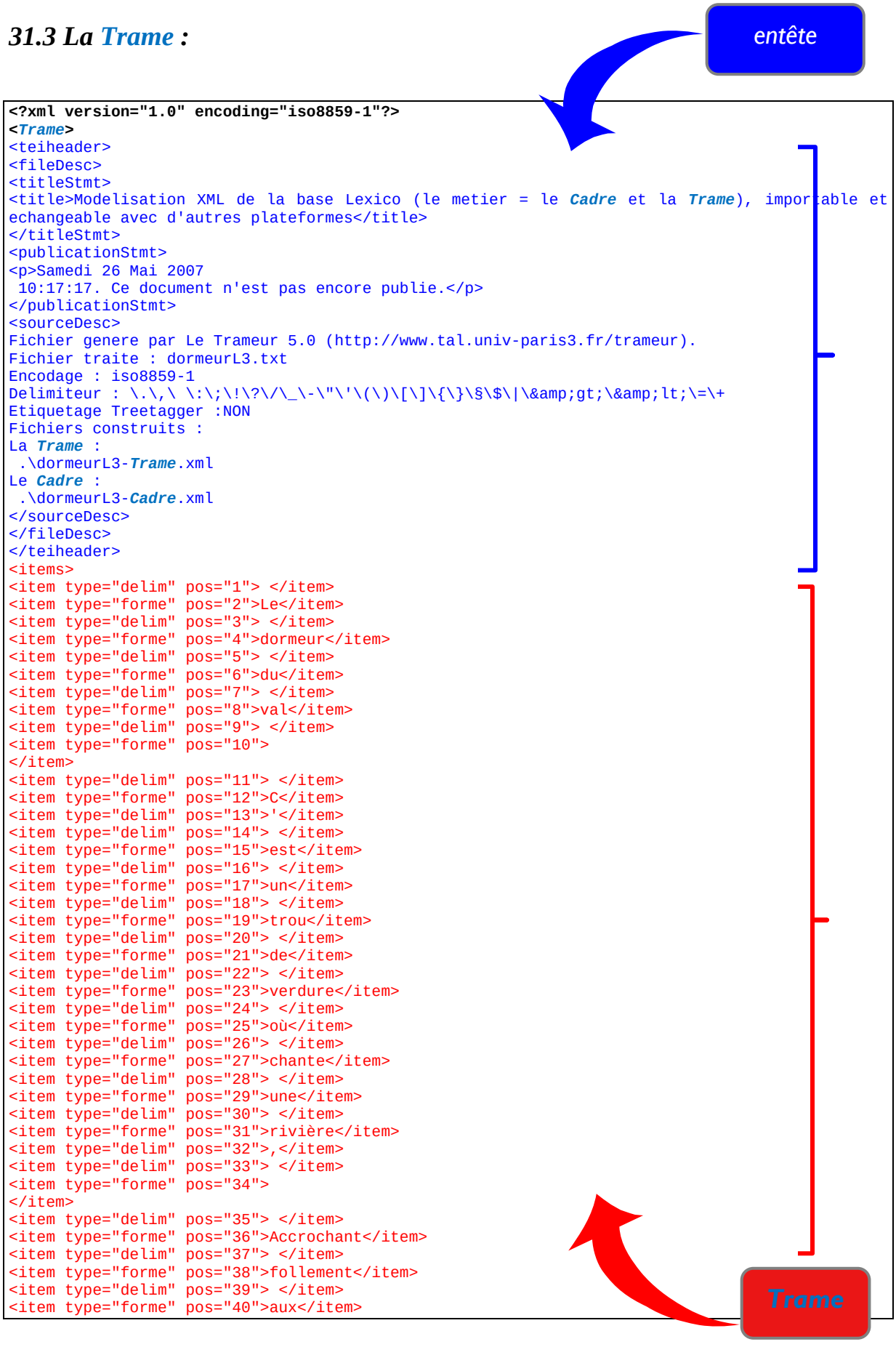
```
<STRUCTURE="TITRE"> Le dormeur du val
<STRUCTURE="TEXTEPOEME">
<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
<LIGNE="VERS2"> Accrochant follement aux herbes des haillons
<LIGNE="VERS3"> D' argent ; où le soleil ; de la montagne fière,
<LIGNE="VERS4"> Luit : c' est un petit val qui mousse de rayons.
```

31.2 Le Cadre :




```
<?xml version="1.0" encoding="iso8859-1"?>
<Cadre>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame),
importable et echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
<p>Samedi 26 Mai 2007
 10:17:17. Ce document n'est pas encore publie.</p>
</publicationStmt>
<sourceDesc>
Fichier genere par Le Trameur 5.0 (http://www.tal.univ-paris3.fr/trameur).
Fichier traite : dormeurL3.txt
Encodage : iso8859-1
Delimiteur : \.\,\ \:\;\!\?\/\_\-\"\'\(\)\[\]\{\}\§\$\|\&amp;gt;\&amp;lt;\=\+
Etiquetage Treetagger :NON
Fichiers construits :
La Trame :
 .\dormeurL3-Trame.xml
Le Cadre :
 .\dormeurL3-Cadre.xml
</sourceDesc>
</fileDesc>
</teiheader>
<acces>
<partition nom="structure">
<partie nom="TITRE" ideb="1" ifin="11"/>
<partie nom="TEXTEPOEME" ideb="11" ifin="100"/>
</partition>
<partition nom="LIGNE">
<partie nom="VERS1" ideb="11" ifin="35"/>
<partie nom="VERS2" ideb="35" ifin="49"/>
<partie nom="VERS3" ideb="49" ifin="75"/>
<partie nom="VERS4" ideb="75" ifin="100"/>
</partition>
</acces>
</Cadre>
```

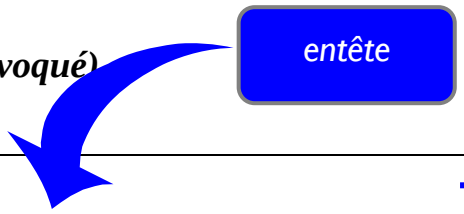
31.3 La Trame :



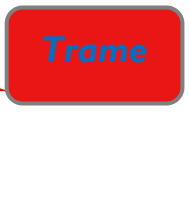

```
<?xml version="1.0" encoding="iso8859-1"?>
<Trame>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame), importable et
echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
<p>Samedi 26 Mai 2007
 10:17:17. Ce document n'est pas encore publie.</p>
</publicationStmt>
<sourceDesc>
Fichier genere par Le Trameur 5.0 (http://www.tal.univ-paris3.fr/trameur).
Fichier traite : dormeurL3.txt
Encodage : iso8859-1
Delimiteur : \.\,\ \:\;\!\?\/\_\-\"\'\(\)\[\]\{\}\§\$\|\&amp;gt;\&amp;lt;\=\+
Etiquetage Treetagger :NON
Fichiers construits :
La Trame :
 .\dormeurL3-Trame.xml
Le Cadre :
 .\dormeurL3-Cadre.xml
</sourceDesc>
</fileDesc>
</teiheader>
<items>
<item type="delim" pos="1"> </item>
<item type="forme" pos="2">Le</item>
<item type="delim" pos="3"> </item>
<item type="forme" pos="4">dormeur</item>
<item type="delim" pos="5"> </item>
<item type="forme" pos="6">du</item>
<item type="delim" pos="7"> </item>
<item type="forme" pos="8">val</item>
<item type="delim" pos="9"> </item>
<item type="forme" pos="10">
</item>
<item type="delim" pos="11"> </item>
<item type="forme" pos="12">C</item>
<item type="delim" pos="13">'</item>
<item type="delim" pos="14"> </item>
<item type="forme" pos="15">est</item>
<item type="delim" pos="16"> </item>
<item type="forme" pos="17">un</item>
<item type="delim" pos="18"> </item>
<item type="forme" pos="19">trou</item>
<item type="delim" pos="20"> </item>
<item type="forme" pos="21">de</item>
<item type="delim" pos="22"> </item>
<item type="forme" pos="23">verdure</item>
<item type="delim" pos="24"> </item>
<item type="forme" pos="25">où</item>
<item type="delim" pos="26"> </item>
<item type="forme" pos="27">chante</item>
<item type="delim" pos="28"> </item>
<item type="forme" pos="29">une</item>
<item type="delim" pos="30"> </item>
<item type="forme" pos="31">rivière</item>
<item type="delim" pos="32">,</item>
<item type="delim" pos="33"> </item>
<item type="forme" pos="34">
</item>
<item type="delim" pos="35"> </item>
<item type="forme" pos="36">Accrochant</item>
<item type="delim" pos="37"> </item>
<item type="forme" pos="38">follement</item>
<item type="delim" pos="39"> </item>
<item type="forme" pos="40">aux</item>
```

```
<item type="delim" pos="41"> </item>
<item type="forme" pos="42">herbes</item>
<item type="delim" pos="43"> </item>
<item type="forme" pos="44">des</item>
<item type="delim" pos="45"> </item>
<item type="forme" pos="46">haillons</item>
<item type="delim" pos="47"> </item>
<item type="forme" pos="48">
</item>
<item type="delim" pos="49"> </item>
<item type="forme" pos="50">D</item>
<item type="delim" pos="51">'</item>
<item type="delim" pos="52"> </item>
<item type="forme" pos="53">argent</item>
<item type="delim" pos="54"> </item>
<item type="delim" pos="55">;</item>
<item type="delim" pos="56"> </item>
<item type="forme" pos="57">où</item>
<item type="delim" pos="58"> </item>
<item type="forme" pos="59">le</item>
<item type="delim" pos="60"> </item>
<item type="forme" pos="61">soleil</item>
<item type="delim" pos="62"> </item>
<item type="delim" pos="63">;</item>
<item type="delim" pos="64"> </item>
<item type="forme" pos="65">de</item>
<item type="delim" pos="66"> </item>
<item type="forme" pos="67">la</item>
<item type="delim" pos="68"> </item>
<item type="forme" pos="69">montagne</item>
<item type="delim" pos="70"> </item>
<item type="forme" pos="71">fière</item>
<item type="delim" pos="72">,</item>
<item type="delim" pos="73"> </item>
<item type="forme" pos="74">
</item>
<item type="delim" pos="75"> </item>
<item type="forme" pos="76">Luit</item>
<item type="delim" pos="77"> </item>
<item type="delim" pos="78">:</item>
<item type="delim" pos="79"> </item>
<item type="forme" pos="80">c</item>
<item type="delim" pos="81">'</item>
<item type="delim" pos="82"> </item>
<item type="forme" pos="83">est</item>
<item type="delim" pos="84"> </item>
<item type="forme" pos="85">un</item>
<item type="delim" pos="86"> </item>
<item type="forme" pos="87">petit</item>
<item type="delim" pos="88"> </item>
<item type="forme" pos="89">val</item>
<item type="delim" pos="90"> </item>
<item type="forme" pos="91">qui</item>
<item type="delim" pos="92"> </item>
<item type="forme" pos="93">mousse</item>
<item type="delim" pos="94"> </item>
<item type="forme" pos="95">de</item>
<item type="delim" pos="96"> </item>
<item type="forme" pos="97">rayons</item>
<item type="delim" pos="98">.</item>
<item type="delim" pos="99"> </item>
</items>
</Trame>
```

31.4 La *Trame* « annotée » (si treetagger est invoqué)




```
<?xml version="1.0" encoding="iso8859-1"?>
<treetagger>
<teiheader>
<fileDesc>
<titleStmt>
<title>Modelisation XML de la base Lexico (le metier = le Cadre et la Trame), importable et
echangeable avec d'autres plateformes</title>
</titleStmt>
<publicationStmt>
<p>Samedi 26 Mai 2007
 10:20:34. Ce document n'est pas encore publie.</p>
</publicationStmt>
<sourceDesc>
Fichier genere par Le Trameur 5.0 (http://www.tal.univ-paris3.fr/trameur).
Fichier traite : dormeurL3.txt
Encodage : iso8859-1
Delimiteur : \.\,\ \:\;\!\?\/\_\-\"\'\(\)\[\]\{\}\§\$\|\&amp;gt;\&amp;lt;\=\+
Etiquetage Treetagger :OUI
Fichiers construits :
La Trame :
 .\dormeurL3-Trame.xml
Le Cadre :
 .\dormeurL3-Cadre.xml
Le fichier a treetagger :
 .\dormeurL3-2-treetagger.txt
Le fichier issu de treetagger :
 .\dormeurL3-etiquette.txt
Le fichier issu de treetagger reconstruisant la Trame initiale :
 .\dormeurL3-Trame-treetager.xml
</sourceDesc>
</fileDesc>
</teiheader>
<items>
<item type="delim" pos="1"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="2"><f>Le</f><c>ADJ</c><l>Le</l></item>
<item type="delim" pos="3"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="4"><f>dormeur</f><c>NOM</c><l>dormeur</l></item>
<item type="delim" pos="5"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="6"><f>du</f><c>PRP_det</c><l>du</l></item>
<item type="delim" pos="7"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="8"><f>val</f><c>NOM</c><l>val</l></item>
<item type="delim" pos="9"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="10"><f>
</f><c>DELIM</c><l>RETURN</l></item>
<item type="delim" pos="11"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="12"><f>C</f><c>NUM</c><l>C</l></item>
<item type="delim" pos="13"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="delim" pos="14"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="15"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="delim" pos="16"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="17"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="delim" pos="18"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="19"><f>trou</f><c>NOM</c><l>trou</l></item>
<item type="delim" pos="20"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="21"><f>de</f><c>PRP</c><l>de</l></item>
<item type="delim" pos="22"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="23"><f>verdure</f><c>NOM</c><l>verdure</l></item>
<item type="delim" pos="24"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="25"><f>où</f><c>PRO_REL</c><l>où</l></item>
<item type="delim" pos="26"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="27"><f>chante</f><c>VER_pres</c><l>chanter</l></item>
<item type="delim" pos="28"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="29"><f>une</f><c>DET_ART</c><l>un</l></item>
<item type="delim" pos="30"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="31"><f>rivière</f><c>NOM</c><l>rivière</l></item>
<item type="delim" pos="32"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="33"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="34"><f>
</f><c>DELIM</c><l>RETURN</l></item>
<item type="delim" pos="35"><f> </f><c>DELIM</c><l>BLANK</l></item>
```



```
<item type="forme" pos="36"><f>Accrochant</f><c>ABR</c><l>Accrochant</l></item>
<item type="delim" pos="37"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="38"><f>follement</f><c>ADV</c><l>follement</l></item>
<item type="delim" pos="39"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="40"><f>aux</f><c>PRP_det</c><l>au</l></item>
<item type="delim" pos="41"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="42"><f>herbes</f><c>NOM</c><l>herbe</l></item>
<item type="delim" pos="43"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="44"><f>des</f><c>PRP_det</c><l>du</l></item>
<item type="delim" pos="45"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="46"><f>haillons</f><c>NOM</c><l>haillon</l></item>
<item type="delim" pos="47"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="48"><f>
</f><c>DELIM</c><l>RETURN</l></item>
<item type="delim" pos="49"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="50"><f>D</f><c>ABR</c><l>D</l></item>
<item type="delim" pos="51"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="delim" pos="52"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="53"><f>argent</f><c>NOM</c><l>argent</l></item>
<item type="delim" pos="54"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="delim" pos="55"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="delim" pos="56"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="57"><f>où</f><c>PRO_REL</c><l>où</l></item>
<item type="delim" pos="58"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="59"><f>le</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="60"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="61"><f>soleil</f><c>NOM</c><l>soleil</l></item>
<item type="delim" pos="62"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="delim" pos="63"><f>;</f><c>DELIM</c><l>;</l></item>
<item type="delim" pos="64"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="65"><f>de</f><c>PRP</c><l>de</l></item>
<item type="delim" pos="66"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="67"><f>la</f><c>DET_ART</c><l>le</l></item>
<item type="delim" pos="68"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="69"><f>montagne</f><c>NOM</c><l>montagne</l></item>
<item type="delim" pos="70"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="71"><f>fière</f><c>ADJ</c><l>fier</l></item>
<item type="delim" pos="72"><f>,</f><c>DELIM</c><l>,</l></item>
<item type="delim" pos="73"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="74"><f>
</f><c>DELIM</c><l>RETURN</l></item>
<item type="delim" pos="75"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="76"><f>Luit</f><c>VER_pres</c><l>luire</l></item>
<item type="delim" pos="77"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="delim" pos="78"><f>:</f><c>DELIM</c><l>:</l></item>
<item type="delim" pos="79"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="80"><f>c</f><c>NOM</c><l>c</l></item>
<item type="delim" pos="81"><f>'</f><c>DELIM</c><l>'</l></item>
<item type="delim" pos="82"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="83"><f>est</f><c>VER_pres</c><l>être</l></item>
<item type="delim" pos="84"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="85"><f>un</f><c>DET_ART</c><l>un</l></item>
<item type="delim" pos="86"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="87"><f>petit</f><c>ADJ</c><l>petit</l></item>
<item type="delim" pos="88"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="89"><f>val</f><c>NOM</c><l>val</l></item>
<item type="delim" pos="90"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="91"><f>qui</f><c>PRO_REL</c><l>qui</l></item>
<item type="delim" pos="92"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="93"><f>mousse</f><c>NOM</c><l>mousse</l></item>
<item type="delim" pos="94"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="95"><f>de</f><c>PRP</c><l>de</l></item>
<item type="delim" pos="96"><f> </f><c>DELIM</c><l>BLANK</l></item>
<item type="forme" pos="97"><f>rayons</f><c>NOM</c><l>rayon</l></item>
<item type="delim" pos="98"><f>.</f><c>DELIM</c><l>.</l></item>
<item type="delim" pos="99"><f> </f><c>DELIM</c><l>BLANK</l></item>
</items>
</treetagger>
```

32 Bibliographie

[Daoust, 2006], Daoust F., «Logiciels d'analyse textuelle : vers un format XML-TEI pour l'échange de corpus annotés» *in Actes des 8èmes Journées d'analyse statistique des données textuelles*, Besançon, 2006. (Disponible en ligne)

[Fleury, Zimina, 2014], Fleury S. and Zimina M. *Trameur: A Framework for Annotated Text Corpora Exploration*, Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations, August 2014, Dublin, Ireland, pages 57-61, (PDF)

[Fleury, 2013a], Fleury Serge. *Approches textométriques des brouillons*, (PDF), séminaire projet ANR ECRITURES, 14 janvier 2013.

[Fleury, 2013b], Fleury Serge. *Le Trameur. Propositions de description et d'implémentation des objets textométriques*, (PDF), (texte en cours).

[Habert & al., 1997], Habert B, Nazarenko A., Salem A. *Linguistiques de corpus*, Paris, Armand-Colin, 1997

[Habert, 2005], Habert B. *Instruments et ressources électroniques pour le français*, Paris, Ophrys, 2005

[Heiden, 2006], Heiden S. « Modèles de données et formats d'échange pour l'interopérabilité des outils de textométrie », in *Actes des 8èmes Journées d'analyse statistique des données textuelles*, Besançon, 2006. (Disponible en ligne)

[Lafon, 1984] Lafon Pierre. *Dépouillements et statistiques en lexicométrie*, Slatkine-Champion, Paris, 1984.

[Lamalle & al., 2006], Lamalle C., Fleury S., Salem A. « Vers une description formelle des traitements textométriques », *in Actes des 8èmes Journées d'analyse statistique des données textuelles*, Besançon, 2006. (Disponible en ligne)

[Lamalle, Salem, 2002], Lamalle C., Salem A. « Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels », in *Actes des 6èmes Journées d'analyse statistique des données textuelles*, St Malo, Inria, 2002. (Disponible en ligne)

[Lebart, Salem, 1994] Lebart L., Salem A. *Statistique textuelle*, Paris, Dunod, 1994. (Disponible en ligne)

[Martinez, 2002] William Martinez, Zimina Maria. "*Utilisation de la méthode des cooccurrences pour l'alignement des mots de textes bilingues*", in Actes JADT'2002[29], Journées

[29] http://www.jadt.org/

Internationales d'Analyse Statistiques des Données Textuelles, St Malo. (Disponible en ligne)

[Martinez, 2003] William Martinez. *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*. Thèse de Doctorat en Sciences du Langage, Université de la Sorbonne nouvelle - Paris 3, sous la direction d'André Salem, Paris. (Disponible en ligne)

[Martinez, Leblanc, 2006] Martinez W., Leblanc J-M. "*L'analyse contrastive des réseaux de cooccurrence Le monde dans les discours des présidents de la Cinquième République*", in Actes JADT'2006, Journées Internationales d'Analyse Statistiques des Données Textuelles, Besançon. (Disponible en ligne)

[Pincemin, 2008], Pincemin Bénédicte (CNRS/ICAR). "Modélisation textométrique des textes", in *Actes JADT 2008*, Journées Internationales d'Analyse Statistiques des Données Textuelles, Lyon 2008. (Disponible en ligne)

[Salem, 1987], Salem André. *Pratique des segments répétés*, Publications de l'INaLF, collection "St.Cloud", Paris, Klincksieck, 1987.

[Söze-Duval, 2008], Keyser Söze-Duval. *Pour une textométrie opérationnelle*. (Disponible en ligne)

33 Glossaire

Sources utilisées pour constituer pour ce glossaire : LEBART Ludovic, SALEM André, *Statistique textuelle*, Paris, Dunod, 1994

Ce glossaire est aussi disponible en ligne :
http://tal.univ-paris3.fr/wakka/wakka.php?wiki=Glossaire

Abréviations :

ac Analyse factorielle des correspondances
acm Analyse des correspondances multiples
cla Classification
sp Méthode des Spécificités
sr Analyse des segments répétés
ling Linguistique
stat Statistique
sa Segmentation automatique

accroissement spécifique (sp)
spécificité calculée pour une partie d'un corpus par rapport à une partie antérieure

analyse factorielle (stat)
famille de méthodes statistiques d'analyse multidimensionnelle, s'appliquant à des tableaux de nombres, qui visent à extraire des "facteurs" résumant approximativement par quelques séries de nombres l'ensemble des informations contenues dans le tableau de départ.

analyse des correspondances (stat)
méthode d'analyse factorielle s'appliquant à l'étude de tableaux à double entrée composés de nombres positifs. L'AC est caractérisée par l'emploi d'une distance (ou métrique) particulière dite distance du chi-2 (ou c2).

caractère (sa)
signe typographique utilisé pour l'encodage du texte sur un support lisible par l'ordinateur.

caractères délimiteurs / non-délimiteurs (sa)
distinction opérée sur l'ensemble des caractères qui entrent dans la composition du texte, permettant aux procédures informatisées de segmenter le texte en occurrences (suite de caractères non-délimiteurs bornée à ses extrémités par des caractères délimiteurs).

On distingue parmi les caractères délimiteurs:

- les caractères délimiteurs d'occurrence (encore appelés "délimiteurs de forme") qui sont en général : le blanc, les signes de ponctuation usuels, les signes de préanalyse éventuellement contenus dans le texte.

- les caractères délimiteurs de séquences : sous-ensemble des délimiteurs d'occurrence correspondant, en général, aux ponctuations faibles et fortes contenues dans la police des caractères.

- les caractères séparateurs de phrase : (sous-ensemble des délimiteurs de séquence) qui correspondent, en général, aux seules ponctuations fortes.

classification (stat)
technique statistique permettant de regrouper des observations ou des individus entre lesquels a été définie une distance.

classification hiérarchique (cla)
technique particulière de classification produisant par agglomération progressive des classes ayant la propriété

d'être, pour deux quelconques d'entre-elles, soit disjointes, soit incluses.

concordance (sa)
l'ensemble de lignes de contexte se rapportant à une même forme-pôle.

contribution absolue (ou contribution) (ac)
contribution apportée par un élément au facteur. Pour un facteur donné, la somme des contributions sur les éléments de chacun des ensembles mis en correspondance est égale à 100.

contribution relative (ou cosinus carré) (ac)
contribution apportée par le facteur à un élément. Pour un élément donné, la somme des contributions relatives sur l'ensemble des facteurs est égale à 1.

cooccurrence (sa) - (une c.)
présence simultanée, mais non forcément contiguë, dans un fragment de texte (séquence, phrase, paragraphe, voisinage d'une occurrence, partie du corpus etc.) des occurrences de deux formes données.

corpus (ling)
ensemble limité des éléments (énoncés) sur lesquels se base l'étude d'un phénomène linguistique.
(lexicométrie) ensemble de textes réunis à des fins de comparaison; servant de base à une étude quantitative.

délimiteurs de séquence (sa)
sous-ensemble des caractères délimiteurs de forme correspondant aux ponctuations faibles et fortes (en général - le point, le point d'interrogation, le point d'exclamation, la virgule, le point-virgule, les deux points, les guillemets, les tirets et les parenthèses).

dendrogramme (cla)
représentation graphique d'un arbre de classification hiérarchique, mettant en évidence l'inclusion progressive des classes.

discours/langue
La langue est un ensemble virtuel qui ne peut être appréhendé que dans son actualisation orale ou écrite; "discours" est un terme commode qui recouvre les deux domaines de cette actualisation.

distance du chi-2
distance entre profils de fréquence utilisée en analyse des correspondances et dans certains algorithmes de classification.

éditions de contextes (sa)
éditions de type concordanciel dans lesquelles les occurrences d'une forme sont accompagnées d'un fragment de contexte pouvant contenir plusieurs lignes de texte autour de la forme-pôle. La longueur de ce contexte est définie en nombre d'occurrences avant et après chaque occurrence de la forme-pôle.

éléments d'un segment (sr)
chacune des formes correspondant aux occurrences qui entrent dans sa composition. ex : A, B, C sont respectivement les premier, deuxième et troisième éléments du segment ABC.

éléments actifs (ac ou acm)
ensemble des éléments servant de base au calcul des axes factoriels, des valeurs propres relatives à ces axes et des coordonnées factorielles.

éléments supplémentaires (ou illustratifs) (ac ou acm)
ensemble des éléments ne participant pas aux calculs des axes factoriels, pour lesquels on calcule des coordonnées factorielles qui auraient été affectées à une forme ayant la même répartition dans le corpus mais participant à l'analyse avec un poids négligeable.

énoncé/énonciation (ling)
à l'intérieur du texte un ensemble de traces qui manifestent l'acte par lequel un auteur a produit ce texte.

facteur (ac ou acm)
variables artificielles construites par les techniques d'analyse factorielle permettant de résumer (de décrire brièvement) les variables actives initiales.

forme (sa) ou forme graphique
archétype correspondant aux occurrences identiques dans un corpus de textes, c'est-à-dire aux occurrences composées strictement des mêmes caractères non-délimiteurs d'occurrence.

forme banale (sp)
pour une partie du corpus donnée, forme ne présentant aucune spécificité (ni positive ni négative) dans cette partie .

forme caractéristique (d'une partie)
synonyme de spécificité positive

forme commune
forme attestée dans chacune des parties du corpus

forme originale (pour une partie du corpus)
forme trouvant toutes ses occurrences dans cette seule partie.

fréquence (sa) (d'une unité textuelle)
le nombre de ses occurrences dans le corpus.

fréquence d'un segment (sr) (ou d'une polyforme)
le nombre des occurrences de ce segment, dans l'ensemble du corpus.

fréquence maximale (sa)
fréquence de la forme la plus fréquente du corpus (en français, le plus souvent, la préposition "de").

fréquence relative (sa)
la fréquence d'une unité textuelle dans le corpus ou dans l'une de ses parties, rapportée à la taille du corpus (resp. de cette partie).

gamme des fréquences (sa)
suite notée Vk, des effectifs correspondant aux formes de fréquence k, lorsque k varie de 1 à la fréquence maximale.

hapax
gr. hapax (legomenon), "chose dite une seule fois".
(sa) forme dont la fréquence est égale à un dans le corpus (hapax du corpus) ou dans une de ses parties (hapax de la partie).

identification (stat, ling, sa)
reconnaissance d'un seul et même élément à travers ses multiples emplois dans des contextes et dans des situations différentes.

index (sa)
liste imprimée constituée à partir d'une réorganisation des formes et des occurrences d'un texte, ayant pour base la forme graphique et permettant de regrouper les références relatives à l'ensemble des occurrences d'une même forme.

index alphabétique (sa)
index dans lequel les formes-pôles* sont classées selon l' ordre lexicographique (celui des dictionnaires).

index hiérarchique (sa)
index dans lequel les formes-pôles sont classées selon l' ordre lexicométrique.

index par parties
ensemble d'index (hiérarchiques ou alphabétiques) réalisés séparément pour chaque partie d'un corpus.

lemmatisation
regroupement sous une forme canonique (en général à partir d'un dictionnaire) des occurrences du texte. En français, ce regroupement se pratique en général de la manière suivante :

- les formes verbales à l'infinitif,
- les substantifs au singulier,
- les adjectifs au masculin singulier,
- les formes élidées à la forme sans élision.

lexical (ling)
qui concerne le lexique ou le vocabulaire.

lexicométrie
ensemble de méthodes permettant d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le vocabulaire d'un corpus de textes.

lexique (ling)
ensemble virtuel des mots d'une langue.

longueur (sa) (d'un corpus, d'une partie de ce corpus, d'un fragment de texte, d'une tranche, d'un segment, etc.) le nombre des occurrences contenues dans ce corpus (resp. : partie, fragment, etc.). Synonyme : taille.

On note: T la longueur du corpus; t j celle de la partie (ou tranche) numéro j du corpus.

longueur d'un segment (sr)
le nombre des occurrences entrant dans la composition de ce segment.

occurrence (sa)
suite de caractères non-délimiteurs bornée à ses extrémités par deux caractères délimiteurs de forme.

ordre lexicographique

- pour les formes graphiques :
  - l'ordre selon lequel les formes sont classées dans un dictionnaire.
  - NB : Les lettres comportant des signes diacrisés sont classées au même niveau que les mêmes caractères non diacrisés, le signe diacritique n'intervenant que dans les cas d'homographie complète. Dans les dictionnaires, on trouve par exemple rangées dans cet ordre les formes : mais, maïs, maison, maître .
- pour les polyformes:
  - ordre résultant d'un tri des polyformes par ordre lexicographique sur la première composante. Les polyformes commençant par une même forme graphique sont départagées par un tri lexicographique sur la seconde, etc.

ordre lexicométrique (sa)

- pour les formes graphiques :
  - ordre résultant d'un tri des formes du corpus par ordre de fréquences décroissantes ; les formes de même fréquence sont classées par ordre lexicographique.
- pour les polyformes:
  - ordre résultant d'un tri par ordre de longueur décroissante des segments, les segments de même longueur sont départagés par leur fréquence, les segments ayant même longueur et même fréquence par l'ordre lexicographique.

paradigme (ling)
ensemble des termes qui peuvent figurer en un point de la chaîne parlée.

paradigmatique (sa)
qui concerne le regroupement en série des unités textuelles, indépendamment de leur ordre de succession dans la chaîne écrite.

partie
(d'un corpus de textes) fragment de texte correspondant aux divisions naturelles de ce corpus ou à un regroupement de ces dernières.

partition

- d'un corpus de textes :
  - division d'un corpus en parties constituées par des fragments de texte consécutifs, n'ayant pas d'intersection commune et dont la réunion est égale au corpus.
- d'un ensemble, d'un échantillon :
  - division d'un ensemble d'individus ou d'observations en classes disjointes dont la réunion est égale à l'ensemble tout entier.

partition longitudinale (sa)
partition d'un corpus en fonction d'une variable qui définit un ordre sur l'ensemble des parties

périodisation (sa)
regroupement des parties naturelles du corpus respectant l'ordre chronologique d'écriture, d'édition ou de parution des textes réunis dans le corpus.

phrase - (sa) fragment de texte compris entre deux séparateurs de phrase.

polyforme (sr)
archétype des occurrences d'un segment; suite de formes non séparées par un séparateur de séquence, qui n'est pas obligatoirement attestée dans le corpus.

ponctuation
Système de signes servant à indiquer les divisions d'un texte et à noter certains rapports syntaxiques et/ou conditions d'énonciation.
(sa) caractère (ou suite de caractères) correspondant à un signe de ponctuation.

pourcentages d'inertie (ac ou acm)
quantités proportionnelles aux valeurs propres dont la somme est égale à 100. Notées ta.

profil (stat et ac) (d'une ligne ou d'une colonne d'un tableau à double entrée)
vecteur constitué par le rapport des effectifs contenus sur cette ligne (resp. colonne) à la somme des effectifs que contient la ligne (resp. la colonne).

répartition (sa) (des occurrences d'une forme dans les parties du corpus)
nombre des parties du corpus dans lesquelles cette forme est attestée.

section (sr)
portion de texte comprise entre deux délimiteurs de section (exemple : le paragraphe, etc.).

segment (sr)
toute suite d'occurrences consécutives dans le corpus et non séparées par un séparateur de séquence est un segment du texte.

segment répété (sr) (ou polyforme répétée)
suite de forme dont la fréquence est supérieure ou égale à 2 dans le corpus.

segmentaire (sr)
ensemble des termes* attestés dans le corpus.

segmentation

opération qui consiste à délimiter des unités minimales dans un texte.

segmentation automatique
ensemble d'opérations réalisées au moyen de procédures informatisées qui aboutissent à découper, selon des règles prédéfinies, un texte stocké sur un support lisible par un ordinateur en unités distinctes que l'on appelle des unités minimales.

séparateurs de phrases (sa)
sous-ensemble des caractères délimiteurs de séquence* correspondant aux seules ponctuations fortes (en général : le point, le point d'interrogation, le point d'exclamation).

séquence (sa)
suite d'occurrences du texte non séparées par un délimiteur* de séquence.

seuil (stat)
quantité arbitrairement fixée au début d'une expérience visant à sélectionner parmi un grand nombre de résultats, ceux pour lesquels les valeurs d'un indice numérique dépassent ce seuil (de fréquence, en probabilité, etc.).

sous-fréquence (sa) (d'une unité textuelle dans une partie, tranche, etc.)
nombre des occurrences de cette unité dans la seule partie (resp. tranche, etc.) du corpus.

sous-segments (sr)
pour un segment donné, tous les segments de longueur inférieure et compris dans ce segment sont des sous-segments. ex : AB et BC sont deux sous-segments du segment ABC.

spécificité chronologique (sp)
spécificité* portant sur un groupe connexe de parties d'un corpus muni d'une partition longitudinale.

spécificité positive (sp)
pour un seuil de spécificité fixé, une forme i et une partie j données, la forme i est dite spécifique positive de la partie j (ou forme caractéristique de cette partie) si sa sous-fréquence est "anormalement élevée" dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou supérieures à la sous-fréquence constatée est inférieure au seuil fixé au départ.

spécificité négative (sp)
pour un seuil de spécificité fixé, une forme i et une partie j données, la forme i est dite spécifique négative de la partie j si sa sous-fréquence est anormalement faible dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou inférieures à la sous-fréquence constatée est inférieure au seuil fixé au départ.

stock distributionnel du vocabulaire (d'un fragment de texte)
le vocabulaire* de ce fragment assorti de comptages de fréquence pour chacune des formes entrant dans sa composition.

syntagmatique (sa)
qui concerne le regroupement des unités textuelles, selon leur ordre de succession dans la chaîne écrite.

syntagme (ling)
groupe de mots en séquence formant une unité à l'intérieur de la phrase.

tableau de contingence (stat)
synonyme de tableau de fréquences ou de tableau croisé: tableau dont les lignes et les colonnes représentent respectivement les modalités de deux questions (ou deux variables nominales) , et dont le terme général représente le nombre d'individus correspondant à chaque couple de modalités.

tableau lexical entier (TLE)
tableau à double entrée dont les lignes sont constituées par les ventilations des différentes formes dans les parties du corpus. Le terme générique k(i,j) du TLE est égal au nombre de fois que la forme i est attestée dans la partie j du corpus. Les lignes du TLE sont triées selon l' ordre lexicométrique des formes correspondantes.

tableau des segments répétés (TSR)
tableau à double entrée dont les lignes sont constituées par les ventilations des segments répétés dans les parties du corpus. Les lignes du TSR sont triées selon l'ordre lexicométrique* des segments. (i.e. longueur décroissante, fréquence décroissante, ordre lexicographique).

tableau lexical
tableau à double entrée résultant du TLE par suppression de certaines lignes (par exemple celles qui correspondent à des formes dont la fréquence est inférieure à un seuil donné).

taille (sa) (d'un corpus)
sa longueur* mesurée en occurrences (de formes simples).

terme (sr)
nom générique s'appliquant à la fois aux formes et aux polyformes. Dans le premier cas on parlera de termes de longueur 1. Les polyformes sont des termes de longueur 2,3, etc.

termes contraints / termes libres
un terme S1 est contraint dans un autre terme S2 de longueur supérieure si toutes ses occurrences sont des sous-segments de segments correspondant à des occurrences du segment S2. Si au contraire un terme possède plusieurs expansions distinctes, qui ne sont pas forcément récurrentes, c'est un terme libre.

types généralisés (Tgens)
unités de dépouillement définies par l'utilisateur à l'aide d'outils permettant d'effectuer automatiquement des regroupements d'occurrences du texte (ex : les occurrences des formes qui commencent par la séquence de caractère patr : patrie, patriotes, patriotisme, etc.).

unités minimales (pour un type de segmentation)
unités que l'on ne décompose pas en unités plus petites pouvant entrer dans leur composition (ex : dans la segmentation en formes graphiques les formes ne sont pas décomposées en fonction des caractères qui les composent)

valeur modale (stat)
valeur pour laquelle une distribution atteint son maximum.

valeurs propres (ac ou acm)
quantités permettant de juger de l'importance des facteurs successifs de la décomposition factorielle. La valeur propre notée la. mesure la dispersion des éléments sur l' axe.a.

valeurs-tests (ac ou acm)
quantités permettant d'apprécier la signification de la position d'un élément supplémentaire (ou illustratif) sur une axe factoriel. Brièvement, si une valeur test dépasse 2 en valeur absolue, il y a 95 chances sur 100 que la position de l'élément correspondant ne puisse être due au hasard.

variables actives
variables utilisées pour dresser une typologie, soit par analyse factorielle, soit par classification. Les typologies dépendent du choix et des poids des variables actives, qui doivent de ce fait constituer un ensemble homogène.

variables supplémentaires (ou illustratives)
variables utilisées a posteriori pour illustrer des plans factoriels ou des classes. Une variable supplémentaire peut-être considérée comme une variable active munie d'un poids nul.

variables de type T
variable dont la fréquence est à peu près proportionnelle à l'allongement du texte. (ex : la fréquence maximale)

variables de type V
variable dont l'accroissement a tendance à diminuer avec l'allongement du texte (ex : le nombre des formes, le nombre des hapax).

ventilation (sa) (des occurrences d'une unité dans les parties du corpus)
La suite des n nombres (n = nombre de parties du corpus) constituée par la succession des sous-fréquences* de

cette unité dans chacune des parties, prises dans l'ordre des parties.

vocabulaire (sa)
ensemble des formes attestées dans un corpus de textes.

vocabulaire commun (sa)
l'ensemble des formes attestées dans chacune des parties du corpus.

vocabulaire de base (sp)
ensemble des formes du corpus ne présentant, pour un seuil fixé, aucune spécificité (négative ou positive) dans aucune des parties , (i.e. l'ensemble des formes qui sont "banales" pour chacune des parties du corpus).

vocabulaire original (sp) (pour une partie du corpus)
l'ensemble des formes originales pour cette partie.

voisinage d'une occurrence (sa)
pour une occurrence donnée du texte, tout segment (suite d'occurrences consécutives, non séparées par un délimiteur de séquence) contenant cette occurrence.

Le Trameur

CLA2T / SYLED - Université Sorbonne nouvelle Paris 3

http://www.tal.univ-paris3.fr/trameur/